DESIGN OF EXPERIMENTS FOR INDUSTRIAL ENGINEERS

by

Virgil L. Anderson
Purdue University

Robert A. McLean
University of Tennessee

Technical Report #82-14

Department of Statistics
Purdue University
May 1982

DESIGN OF EXPERIMENTS FOR INDUSTRIAL ENGINEERS

Professor Virgil L. Anderson
Department of Statistics
Purdue University
W. Lafayette, Indiana 47907
(317) 494-6033

and

Professor Robert A. McLean
Department of Statistics
University of Tennessee
Knoxville, Tennessee 37916
(615) 974-2556

Published in: Handbook of Industrial Engineering
1982, John Wiley & Sons, Inc.
Editor: Dr. G. Salvendy

I. OVERVIEW

Why should Industrial Engineers be interested in designing experiments? They have looked at data from so called "experiments" for years. In some cases the results have been confusing or the data points looked wrong when compared to theory or preconceived ideas. In such cases people have been known to rerun the experiment, forget the experiment completely, or even change some data points in order to make the results look better.

A comment we statisticians hear so often is that designing experiments takes so much time and the engineer cannot afford to take more data. In many instances data, often taken haphazardly, are already available and it is not understood why the statistician cannot just analyze those data and interpret the results, rather than take more data from a well-designed experiment.

The main reason for taking data from a designed experiment is that the investigator can place a given probability statement on the results, if the experiment has been <u>carefully</u> designed. Also, engineers who use data to help them draw conclusions usually want to know how widely the results will apply (Inference Space).

In almost all cases dealing with data, the experimenter wants to keep the number of observations small. Carefully designed experiments will allow minimum sample size for a specified problem if the variation is known. If the variation is not known (as in almost all cases), a small sample or a pre-experiment may be used to estimate the variation before the overall experiment is run.

Is all of this magic? No! It requires thinking, co-operation, work and a willingness to learn basic concepts such as "confounding" and "biasedness".

Let us look at the past a bit and then turn our attention to a couple of simple design examples before delving into a few good industrial engineering design of experiments to complete this chapter. The reader must understand that to run experiments efficiently, he must read and study books such as Anderson and McLean (1974) thoroughly.

For centuries "man" has run experiments to answer questions. The idea of taking a sample to draw conclusions about a much larger group (population) is not new. Cochran and Cox (1957), however, point out that randomization is a relatively new concept and Anderson and McLean (1974) indicate that recently too many investigators have not been careful enough in defining how wide the results of their experiments apply (Inference Space).

To give direction in the thinking of whether investigators should actually take care to design experiments or not, let us consider the following example:

In a small shop a pattern maker wanted to buy a new lathe. He had narrowed the decision down to two brands, and these two manufacturers offered to let him try the lathes before making the decision which to buy. Company representatives brought the two lathes to his shop, so he set up an experiment to help him decide between the two.

He thought 16 different patterns (requiring approximately the same time to cut) would be enough for him to make a decision if he used the <u>time required to cut a pattern to specifications</u> as the basis for his decision (criterion). Of course this meant that he would buy the lathe which required less time per pattern since the cost of the lathes was equal.

He was a "careful" experimenter and required that each pattern be used on both machines. This would allow him to take the difference in time required to cut the pattern on each machine and make his decision easy.

To begin the experiment he flipped a coin to decide which machine should be used first throughout the experiment. They layout for the experiment is

	Pattern 1	2	3	16
Lathe 1	1	3	5	31
2	2	4	6	32

where the numbers inside the table indicate the order the patterns are to be cut on the lathes.

It has been our experience that many experiments are run this way or they are run without the first randomization because it is easy to keep the records straight as the investigator goes through the experiment. This is not a very thoroughly designed experiment because if the pattern maker learns how to make that given pattern on lathe 1, he will probably retain some of that knowledge when he gets to lathe 2. Hence if it should turn out that he can, in general, cut patterns faster on lathe 2, he will not know for sure whether it was due to the lathe being "better" or his learning from the first cut on lathe 1. This is an example of "confounding". That is, the effect of <u>lathe</u> and <u>learning</u> cannot be separated. Hence there is a "biased" estimate of the effect of lathe.

To improve the design of this simple experiment many people would completely randomize the order of cutting the patterns. One possible layout of the completely randomized designed experiment is:

Pattern

Lathe	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	27	9	7	15	24	19	10	14	13	8	29	12	21	11	28	17
2	3	25	2	16	22	1	20	31	23	5	26	4	30	18	32	6

where the numbers inside the table indicate the order the patterns are to be cut on the particular lathe. For example, the first cutting would be pattern number 6 on lathe 2 and the last one (32) would be pattern number 15 on lathe 2.

While complete randomization provides a more thoroughly designed experiment, one can easily see that peculiar sequences can be obtained. Notice in this so called "completely" randomized design that lathe 2 is used to cut the first six times and lathe 1 is used for the next nine. It is not known whether this sequencing interferes with the correct decision to buy the better lathe or not, but it is known (mathematically) that complete randomization does provide unbiased estimates of the effects.

Another way to run this experiment (the best way, we think) is to use the layout

	Sequence 1 Patterns								Sequence 2 Patterns							
Cut	4	7	15	6	3	14	1	10	16	2	9	5	12	13	8	11
First	Lathe 2								Lathe 1							
Second	Lathe 1								Lathe 2							

The operational procedure for this approach is more complicated and the need for the additional detail is hard to explain to some experimenters. It is necessary to run the same number of patterns first on lathe 1 as on lathe 2 in order to obtain an unbiased estimate of the difference in time required for the two lathes. We insist on randomly assigning the various patterns to the two different sequences so that one lathe will not be favored over the other as the result of unsuspected differences among the patterns. One possible selection is the use of patterns 4, 7, 15, 6, 3, 14, 1, and 10 for Sequence 1 as is shown in the above layout. One additional precaution must be taken in order to guard against such effects as fatigue. This can be accomplished by randomly selecting the order in which the patterns are actually cut. This can be done by randomly drawing the numbers 1 through 16. One such sequence would be 14, 11, 6, ..., 10. Thus pattern 14 would be cut on lathe 2 and then on lathe 1, this would be followed by pattern 11 first on lathe 1 and then on lathe 2, and etc. (not necessarily alternating).

This last design is discussed in detail later and is called a "crossover" design.

Another example of a designed experiment (this one ill-defined) occurred a number of years ago in a large manufacturing company. A man working in the production area set up an experiment to test a new alloy, possibly one to replace an old one in production. He ran only one heat (batch) of metal with the new alloy and another heat with the old one. Taking one ingot from each heat and 30 pieces of metal from each ingot he proceeded to test each of the 60 pieces for the property in which he was interested. With the data he made a one-way analysis of variance (ANOVA) on the alloys using the pieces within ingots with 58 degrees of freedom as the error. The results showed that the new alloy was "better" than the old one, and the experimenter convinced the vice president in charge of production to change the production procedures so that the new alloy would be used in the future. Since the experimenter had used a "designed experiment" and had tested the data "statistically", the vice president concluded there could be no doubt that the new one was better.

The change cost the company $200,000, and after 2 years in the field there was as much trouble with the product made from the new alloy as there had been with the old product. The vice president was disgusted and called one of the authors of this chapter to say he would never allow his company to use designed experiments again. After some discussion, the vice president allowed the author to talk with the experimenter to find out how the experiment was conducted.

In wanting to keep the cost of the experiment low, the experimenter did not consider the possibility that the property in which he as interested varied considerably both from heat to heat and from ingot to ingot within a heat. From an ANOVA point of view, his expected means squares (EMS) should have been what is shown as follows:

ANOVA of Alloy Problem

Source	df	EMS
Alloys	1	$\sigma_p^2 + 30\sigma_I^2 + 30\sigma_H^2 + 30\phi(A)$
Pieces in ingots	58	σ_p^2
Total	59	

Hence, rather than testing that the alloy effect was zero [$\phi(A)=0$], he was really testing that the total effect for ingots, heats, and alloy was ZERO[$30\sigma_I^2+30\sigma_H^2+30\phi(A)=0$]. Since the long-run production of the new alloy did not produce the improvement seen in the experiment, $\phi(A)$ must equal zero. Thus it must have been that σ_I^2 and/or σ_H^2 were/was not zero. The notation utilized here is that ϕ represents a fixed effect, i.e., the effect of two specific alloys, and σ^2 represents a random effect, e.g., the ingots are a random sample from all possible ingots.

When the results were explained, the vice president was willing to use the design of experiments again on this type of problem but insisted upon having more than one heat for each alloy.

Still another example (an excellent one) occurred in a company fabricating men's synthetic felt hats.

The manufacturer had experienced extreme difficulty in producing these hats so that the flocking appeared on the molded rubber base in a uniform fashion to simulate the real felt hats. In order to approach this problem, a committee was formed consisting of a development engineer, a manufacturing foreman, a chief operator, a sales representative, and a statistician. The statistician's job was to obtain from these people all possible causes of imperfect hats. Factors which were thrown out for discussion were as follows: thickness of foam rubber base, pressure of molding, time of molding, viscosity of the latex used to glue the flocking to the molded rubber base, age of the latex, nozzle size of several different spray guns, direction of spraying, condition of the flocking, speed of drying, and the effect of location within the drying furnace.

After considerable discussion the committee finally decided that the most serious problems were probably connected with the nozzle size and the pressure under which the latex was sprayed. In arriving at these various factors the committee essentially forced a review of the entire production process. This in itself led to a better understanding of the production process and an eventual solution to the problem.

In arriving at the most reasonable causes of defective hats, the committee action essentially required a critical review of all production process factors. In this review the chief operatior brought out the standard operating levels of the nozzle size as well as the pressure under which the latex was sprayed. Talk with the chief operator revealed that the latex pressure varied considerably due to the viscosity of latex, and from this information the pressure levels were eventually obtained. Additional inquiry ascertained that the manufacturing area had two different nozzle sizes that had been used interchangeably; consequently, the two sizes became the basis for these factor levels.

The authors feel that one is almost always able to find realistic levels for all major factors through committee action of this type. Occasionally considerable effort is needed to find out just how shoddy one's manufacturing operation really is, and this example shows that actual production operations will allow their process to operate at various levels as long as it works. The determination of the optimal levels is, of course, the desired end of the experimental investigation.

As one can imagine in the manufacturing of synthetic hats, the measuring of product quality is a very difficult task. Consequently, the method that was used in the experiment referred to above was to visually grade the finished hat on the following items: the hungry appearance of the flocking, the starchy appearance of the flocking, and the appearance of brim. During the course of the investigation these responses were found to be essentially independent of each other and consequently could be treated as three separate dependent variables which could be investigated on a one-at-a-time basis. The standards for grading each of these factors were arrived at again through committee action which eventually resulted in a visual display board and gave the inspectors a realistic means of grading each of these dependent variables.

One of the most difficult problems in certain types of industrial experimentation is the specification of a dependent variable. It is usually quite obvious what the variable should be; however, the methods of measurement are sometimes quite difficult. In ideal cases this would be merely the measured value obtained by some simple inspection tool, while in other cases it can be a value that is almost impossible to measure and will have to be graded by one or more inspectors.

Having agreed on the variables to analyze, the committee decided to include six nozzle locations, each with a high and low latex viscosity and each with a high and low air pressure, plus two types of base material as factors and levels in the experiment. This required a 2^{13} factorial type experiment where all possible combinations of the 13 factors were to be considered. After some work a fractional replicated design was run requiring only 256 of the $2^{13} = 8192$ combinations. This design allowed complete information on all main effects and two factor interactions. Fractional factorials are described in detail in Anderson and McLean (1974).

The experiment required certain blocking (described later in this chapter) procedures, and the results were excellent. The company was able to pinpoint the difficulty in the spraying mechanism, make the necessary changes and produce a profitable product in six months. An interesting sidelight to this experiemnt was that the company was losing about $8.00 per hat when the experiment began, and the committee was given a deadline of one year to solve the problem or the company would discontinue the product. With the results from that one large experiment, the company was able to make a profit and a competing company bought the whole process within the year.

This concludes the overview section dealing with examples of designed experiments. Next, in the background section, we attempt to describe our present day basic philosophy that has been successfully used since 1974.

II. BACKGROUND

One approach to designing experiments in the last decade considers three essential ingredients of a well designed experiment expressed in the book by Anderson and McLean (1974). These three ingredients, in order of importance, are:

A. Inference Space

B. Randomization

C. Replication.

Inference space is a phrase to replace the term "population", usually used by statisticians. It has been our experience that the term "inference space" demands more attention from the research worker.

The phrase "inference space" means the limits to which the investigator may use the results of the experiment. Common practice at present is for the research worker to indicate how extensively he wishes the results to apply before the experiment is set up. This requires that he define the experimental units he is to use in his research and that will be the basis for the inferences. He must also define the time interval and the geographical extent to which he wishes the results to apply, and then decide which levels of all factors he wants controlled in the experiment. Ordinarily more time should be spent on this phase of designing the experiment than either of the other two (randomization and replication) because without the inference space clearly defined the best so called "designed experiment" may be worthless to the investigator. We define this ingredient (inference space) as a part of the designed experiment. Hence there can be no "best designed experiment" without a carefully defined inference space.

Randomization is the next most important ingredient in designing experiments. It must be present in the experiment for probability statements to be made. Fisher (1960), p.17, expressed the idea that it is the physical basis of the validity of the test. It is also the basis of validity of confidence intervals.

Included in the ingredient, randomization, is another concept, "restriction" on randomization. To understand restriction on randomization let us first expalin "completely randomized" which means no restriction on randomization with reference to the source of the experimental unit. This concept can be seen by considering a factor, t, with five levels and three experimental units treated with each of the five levels of factor t completely at random.

Assume there are 15 randomly drawn experimental units from the inference space to be used for the entire experiment. One way to obtain a completely randomized design is to select a random number between 1 and 5, say 2. Then the first experimental unit must receive treatment 2 or the second level of factor t. Select another number between 1 and 5, say 5, then the second experimental unit must receive treatment 5. Continue sampling or drawing random numbers in this manner until the 15 esperimental units have all been "treated". This sampling procedure requires that each level of factor t is represented three times in the experiment and the design of this experiment is called "completely randomized".

The mathematical model for analyzing the data from such an experiment is

$$y_{ij} = \mu + T_i + \varepsilon_{(i)j}: \quad i=1,2,\ldots,5; \quad j=1,2,3 \tag{1}$$

where

y_{ij} = the response from experimental unit j treated with level i of factor t,

μ = overall mean

T_i = effect of the i^{th} level of factor t,

$\varepsilon_{(i)j}$ = the experimental error caused by the j^{th} experimental unit nested in the i^{th} level of factor t.

The assumptions for the analysis of the data in a model such as this are:

a) y_{ij} is a random variable.

b) The variances of the responses within levels of factor t are equal.

c) The model is additive.

d) The experimental error is $NID(0,\sigma^2)$, normal and independently distributed with mean zero and variance, σ^2.

This complete randomization assures the experimenter that the experimental units from the inference space have three (3) opportunities for selection for each treatment. These three experimental units for each treatment then allow a measure of the variation within the treatments, which is the representation of the variation across the entire inference space. It follows, then, that the test of significance on treatment effects (T_i in equation (1)) must be based on the excessive amount of variation among the means of the treatments over the amount of variation obtained from within the treatments where the variation within the treatments is accounted for by the variance due to $\varepsilon_{(i)j}$ in equation (1). Hence in equation (1) the $\varepsilon_{(i)j}$ is the correct error for evaluation T_i.

If, however, the sampling procedure was such that the first random draw, level 2, was used on the first three experimental units; then the second draw, level 5, was used on the next three experimental units and so on, we would have a "restriction" on randomization because the randomization procedures were allowed only five times, not 15 as is required for complete randomization. If, as is frequently the case, there tends to be similarity between adjacent units in space or time, the variation within the group of three is smaller than the variation between the groups of three. In this design then, the variation due to treatments is not separable from the variation between groups of three units (treatments confounded with groups). If then there is a source of error variation between groups, it will not be possible to test for treatments, and the following model depicts this:

$$y_{ij} = \mu + T_i + \delta_{(i)} + \varepsilon'_{(i)j} \tag{2}$$

where: $\delta_{(i)}$ is the "restriction" error or that random component due to the i^{th} group of units (note how the subscript is identical to the subscript to T_i indicating complete confounding), and

$$\varepsilon'_{(i)j} \text{ is NID}(0,\sigma^2_{\varepsilon'}),$$

the error due the j^{th} unit in the i^{th} group. It is the variation due to $\delta_{(i)}$ that is representative of the variation across the inference space, whereas the variation due to the $\varepsilon'_{(i)j}$ in equation (2) represents only a small portion across the inference space. Hence one needs an estimate of σ^2_δ to test for the effect of treatments. Of course $\delta_{(i)}$ has no degrees of freedom, which indicates this is a poor design and should not be used.

Using the algorithm for deriving the expected mean squares described in Chapter 2 of Anderson and McLean (1974), we can show that the correct error term for testing teatments is $\delta_{(i)}$. However, there is no estimate of this error in equation (2) unless the whole experiment is repeated. Hence the "restriction" on the randomization has cause $\delta_{(i)}$ which, in turn, tells the experimenter that this sampling procedure is not a good one even before the first observation is made. This, then, allows the experimenter to change his design early, before he has taken data that will not give a good analysis.

Before explaining the third ingredient of designed experiments, namely replication, a few words should be stated about special cases where randomization may not be required. If an investigator can show by actual experimentation that the results are the same whether randomization take place or not it may not be necessary to randomize. This will happen on extremely well controlled experiments only. We are familiar with an example on a one cylinder engine gasoline consumption laboratory experiment in which it did not matter whether speeds were randomized or taken in order. The reason for this was that the controls on speed were so precise and the set-ups so repeatable that the errors were identical within the capabilities of the recording equipment. In terms of the model for analyzing the data from this experiment [similar to equation (2)] $\delta_{(i)}$ is approximately zero.

With this non-random possibility in mind and knowing an experiment cannot be designed well without knowing the inference space, we rank inference space above randomization in importance when considering ingredients of a well designed experiment.

Finally the third ingredient, replication, is quite often required for an estimate of an error term, or to provide the basis for making decisions on the importance of factors contributing to the response variables. In addition, as the number of observations increases on a given treatment the more precise the estimate of the effect of the treatment becomes, or the smaller its variance becomes.

If, however, previous experimentation has shown certain information is available, e.g. the variance is known or that higher order terms in the model are zero, it may not be necessary to replicate the entire experiment. In fact there are many good experiments run with fewer than the total number of levels of the factors in a "factorial" experiment. These experiments are called "fractional replicated factorials".

With the various well-designed experiments without complete replications in mind, the ingredient "replication" is placed in third position behind inference space and randomization.

Before going into designs and analyses it should be understood that the readers of this material should understand (1) basic statistical concepts, (2) distributions such as the mean $\bar{y}$, t, χ^2, F and (3) analysis of variance and regression models.

III. DESIGNS

A definition of a designed experiment is an arrangement of the experimental material, including randomization of experimental units to the treatments so that statistical tests of significance (and confidence intervals) on the effects and interactions of the factors being studied can be made. In order to accomplish this, care must be taken to set up the arrangement efficiently (keep the cost reasonably low) and, at the same time, cover the inference space. For coverage of designs, major headings are used to indicate designs encountered by us in engineering studies.

A. BLOCK DESIGNS

1. Importance of Blocking (Handling Extraneous Variables)

Many authors of books on design of experiments express the importance of "blocking", placing all treatments or all combinations of the levels of all factors of interest in a homogeneous group (thereby removing some of the effect of an extraneous variable from the experimental error) and repeating this group or block in time and/or space with different experimental units. To show this concept we use mathematical equations which are to be used as the basis for analyses of the data from the designed experiment.

Returning to the concepts of setting up equations (1) and (2) it follows that another design of the experiment is to arrange three blocks of five treatments each, where the five experimental units were randomized onto the five treatments per block. It follows that the equation to be used as the basis for the analysis is:

$$y_{ij} = \mu + B_i + \delta_{(i)} + T_j + \varepsilon_{ij} \tag{3}$$

where: Y_{ij} and μ have the same meaning as they did for equation (2)

B_i = effect of block i

$\delta_{(i)}$ is similar to $\delta_{(i)}$ in equation (2)

T_j = effect of the j^{th} treatment

ε_{ij} = error due to the j^{th} treatment in block i (assuming there is no interaction).

Since the experimenter is interested in testing for treatment effects only, this is an excellent design because ε_{ij} is the basis for the test of T_i since the variation across treatments is compared to the variation due to ε_{ij}'s. If it should turn out that the effects of B_i and $\delta_{(i)}$ are zero, B_i, $\delta_{(i)}$ and ε_{ij} may be pooled and equation (3) becomes equation (1). This completes the demonstration that, in general, blocking is always worthwhile in experiments and should be used whenever possible.

Latin Square designs, used correctly, merely extend restriction on randomization in one more dimension. Too many engineers <u>incorrectly</u> use Latin Square designs as special fractional factorials. Reference Chapter 8, Anderson and McLean (1974). Since it is difficult to make the reader understand these points in the space allowed here, we recommend that Latin Squares not be used by the novices.

2. Incorrect Use of Blocks (Treatments Used as Blocks)

In the previous example, the effect of $\delta_{(i)}$, the restriction error caused by blocks, did not decrease the importance of blocking because the experimenter was interested in the effect of blocking plus the restriction error in reducing the estimated experimental error, $\hat{\varepsilon}_{ij}$. In that case, $\hat{\varepsilon}_{ij}$ was the basis for testing treatment effects only.

Now, one must consider the case in which the blocking concept is used incorrectly. Consider an example in evaluating microforms where the interest was in reducing user dissatisfaction. The experimenter wanted to use two types of projection [Front (F), Back (B)] as blocks and randomize four screen angles (0°,45°,90°,105°) twice within each "block". The measured variable was seconds required to read the material presented each time. Pictorially the design was

Types of Projection

Front	Back
90°	45°
0°	105°
45°	0°
90°	105°
105°	90°
0°	45°
45°	0°
105°	90°

The correct degrees of freedom and model for the analysis of the data from this design is:

$$16 = 1 + 1 + 0 + 3 + 3 + 8$$

$$y_{ijk} = \mu + T_i + \delta_{(i)} + A_j + TA_{ij} + \varepsilon_{(ij)k} \qquad (4)$$

$$i = 1,2; \quad j = 1,2,3,4; \quad k = 1,2$$

where y_{ijk} = seconds required to read the microform from the k^{th} observation using the j^{th} angle and the i^{th} type of projection,

μ = overall mean

T_i = effect of the i^{th} type of projection (fixed)

$\delta_{(i)}$ = restriction error due to all of the angles used with the i^{th} type of projection before the other type of projection is used.

A_j = effect of the j^{th} angle (fixed)

TA_{ij} = effect of the interaction of the i^{th} type with the j^{th} angle,

$\varepsilon_{(ij)k}$ = error due to the k^{th} observation within the i^{th} type j^{th} angle, assumed NID$(0,\sigma^2)$.

The analysis of variance (ANOVA) appropriate to analyze data using equation (4) is the following:

ANOVA
[Based on equation (4)]

Source	df	EMS
Types of Projection (T_i)	1	$\sigma_\varepsilon^2 + 8\sigma_\delta^2 + 8\phi(T)$
Screen Angles (A_j)	3	$\sigma_\varepsilon^2 + 4\phi(A)$
TA_{ij}	3	$\sigma_\varepsilon^2 + 2\phi(TA)$
Repeats within (T-A)	8	σ_ε^2
Combination $\varepsilon_{(ij)k}$		

It is apparent from the EMS (expected mean squares) column above that there is no test for types of projection because there is no source with an EMS of $\sigma_\varepsilon^2 + 8\sigma_\delta^2$. Hence this is an incorrect use of blocks. In order to obtain the source to estimate $\sigma_\varepsilon^2 + 8\sigma_\delta^2$ there must be a replicate of the experiment. This concept is demonstrated in the following section.

3. CORRECT USE OF BLOCKS

The example used in this section is a continuation of the example used in the previous section.

To make the experiment a good one, the investigator should set up the eight treatment combinations as follows (one possible randomization):

Blocks

1			2	
F	45°		B	105°
B	0°		F	90°
B	105°		B	45°
F	0°		F	105°
B	90°		F	45°
F	105°		B	0°
B	45°		B	90°
F	90°		F	0°

The degrees of freedom and the model for the analysis of the data from this design is as follows:

$$30 = 1 + 1 + 0 + 1 + 1 + 3 + 3$$

$$y_{ijk} = \mu + B_i + \delta_{(i)} + T_j + BT_{ij} + A_k + BA_{ik}$$

$$+ 3 + 3 + 0$$

$$+ TA_{jk} + BTA_{ijk} + \varepsilon_{(ijk)}$$

where: $i = 1,2$; $j = 1,2$; $k = 1,2,3,4$

B_i = effect of i^{th} block (random),

$\delta_{(i)}$ = restriction error using blocks correctly,

T_j, A_k and TA_{jk} = same as T_i, A_j, TA_{ij} of eq. (4)

BT_{ij},BA_{ik},BTA_{ijk} = appropriate errors for testing (note arrows) and they would probably pool for a common error,

$\varepsilon_{(ijk)}$ = residual error, not estimable with one observation per treatment combination per block.

This is expalined in detail in Anderson and McLean (1974).

B. REPEATED MEASURES AND CROSS OVER DESIGNS

A research worker in an industrial plant dealing with soldering parts on electronic equipment was interested in studying "paced" and "unpaced" production. He set the experiment up in a plant which used women only, and recorded the ages of the women. He grouped the women into ages: Young (18-23), Middle (30-35) and Old (52-57). He was able to obtain 6 women in each of those age groups which allowed him to have a sequence (selected at random) of (paced, unpaced) and (unpaced, paced) for two weeks for three women in each group. The layout or design of the experiment can be portrayed as follows:

Ages

	Young		Middle		Old	
	Sequence		Sequence		Sequence	
	1	2	1	2	1	2
	Women	Women	Women	Women	Women	Women
Weeks	1 2 3	4 5 6	7 8 9	10 11 12	13 14 15	16 17 18
1	P	U	U	P	U	P
2	U	P	P	U	P	U

If the assumptions of the analysis of variance (ANOVA) have been met, and ANOVA of the data (different response variables) from this experiment is:

ANOVA
(CROSS OVER DESIGN)

Source	df
Ages (A)	2
Sequence (S)	1
AS	2
Females in (A-S) cells (F)	12
Weeks (W)	1
Paced vs Unpaced (P)	1
AP	2
SP	1
ASP	2
Residual	11

Refer to Grizzle (1965), Mayers (1979) and Anderson and McLean (1974) and for details of this type design and analysis. For designs and analyses utilizing extensions of these designs to more sequences refer to Albert, et al (1979) and Westlake (1979).

C. OTHER DESIGNS

There are many other designs used in industrial engineering dealing with factorial experiments and fractional factorials plus response surface designs. References for these designs and analyses are Myers (1971), Anderson and McLean (1974) and Box, et al (1978).

IV. REFERENCES

[1] Albert, K. S., Brown, S. W., Jr., DeSanto, K. A., DiSanto, A. R., Stewart, R. D. and Chen, T. T., "Double Latin Square Study to Determine Variability and Relative Bioavailability of Methylprednisolon," Jour. of Pharm. Sci., Vol. 68, No. 10, 1312-1316, October 1979.

[2] Anderson, V. L. and McLean, R. A., "Design of Experiments: A Realistic Approach," Marcel Dekker, N.Y., 1974.

[3] Box, G. E. P., Hunter, W. G. and Hunter, J. S., "Statistics for Experimenters," Wiley, N.Y., 1978.

[4] Cohcran, W. g. and Cox, G. M., "Experimental Designs," Wiley, N.Y., 1957, Second Edition.

[5] Fisher, R. A., "The Design of Experiments," Seventh Edition, Hafner, N.Y., 1960.

[6] Grizzle, J. E., "The Two-Period Change-Over Design and Its Use in Clinical Trials," Biometrics, Vol. 21, p. 467-480, 1965.

[7] Myers, J. L., "Fundamentals of Experimental Design," Allyn and Bacon, Third Edition (1979).

[8] Myers, R. H., "Response Surface Methodology," Allyn and Bacon, 1976.

[9] Westlake, W. J., Statistical Aspects of Comparative Bioavailability Trials, Biometrics, Vol. 35, p. 273-280, 1979.

[10] Winer, B. J., "Statistical Principles in Experimental Design," Second Edition, McGraw Hill, 1971.

82-14

Handbook of Industrial Engineering, Edited by G Salvendy John Wiley & Sons Inc NY (1982)

CHAPTER 13.4

Design of Experiments for Industrial Engineers

Grissom 166

6/25/95 – 1230–31

VIRGIL L. ANDERSON
Purdue University

ROBERT A. McLEAN
University of Tennessee

13.4.1 OVERVIEW

Why should industrial engineers be interested in designing experiments? They have looked at data from so-called experiments for years. In some cases the results have been confusing, or the data looked wrong when compared to theory or preconceived ideas. In such cases people have been known to rerun the experiment, forget the experiment completely, or even change some data in order to make the results look better.

A comment that statisticians hear so often is that designing experiments takes so much time and that the engineer cannot afford to take more data. In many instances data, often taken haphazardly, are already available, and it is not understood why the statistician cannot just analyze those data and interpret the results rather than take more data from a well-designed experiment.

The main reason for taking data from a designed experiment is that the investigator can place a given probability statement on the results if the experiment has been *carefully* designed. Also, engineers who use data to help them draw conclusions usually want to know how widely the results will apply (inference space).

In almost all cases dealing with data, the experimenter wants to keep the number of observations small. Carefully designed experiments will allow minimum sample size for a specified problem if the variation is known. If the variation is not known (as in almost all cases), a small sample or a pre-experiment may be used to estimate the variation before the overall experiment is run.

Is all of this magic? No! It requires thinking, cooperation, work, and a willingness to learn basic concepts such as "confounding" and "biasedness."

Let us look at the past a bit and then turn our attention to a couple of simple design examples before delving into a few good industrial engineering designs of experiments to complete this chapter. The reader must understand that, to run experiments efficiently, he or she must read and study books, such as Anderson and McLean,[1] thoroughly.

For centuries humans have run experiments to answer questions. The idea of taking a sample to draw conclusions about a much larger group (population) is not new. Cochran and Cox,[2] however, point out that randomization is a relatively new concept, and Anderson and McLean[1] indicate that recently too many investigators have not been careful enough in defining how wide the results of their experiments apply (inference space).

Design Examples

A Pattern Maker

To give direction to the thinking of whether investigators should actually take care to design experiments or not, let us consider the following example: In a small shop a pattern maker wanted to buy a new lathe. He had narrowed the decision down to two brands, and these two manufacturers offered to let him try the lathes before deciding which to buy. Company representatives brought the two lathes to his shop, and he set up an experiment to help him decide between the two.

He thought 16 different patterns (requiring approximately the same time to cut) would be enough for him to make a decision if he used the *time required to cut a pattern to specifications* as the basis for his decision (criterion). Of course this meant that he would buy the lathe that required less time per pattern since the cost of the lathes was equal.

He was a "careful" experimenter and required that each pattern be used on both machines. This would allow him to take the difference in time required to cut the pattern on each machine, making his decision easy.

To begin the experiment, he flipped a coin to decide which machine should be used first throughout the experiment. The layout for the experiment was as follows:

	Pattern				
Lathe	1	2	3	....	16
1	1	3	5	....	31
2	2	4	6	....	32

The numbers inside the table indicate the order in which the patterns were to be cut on the lathes.

It has been the authors' experience that many experiments are run this way, or they are run without the first randomization because it is easy to keep the records straight as the investigator goes through the experiment. This is not a very thoroughly designed experiment, because if the pattern maker learns how to make that given pattern on lathe 1, he will probably retain some of that knowledge when he gets to lathe 2. Hence, if it should turn out that he can, in general, cut patterns faster on lathe 2, he will not know for sure whether it was due to the lathe's being "better" or to his learning from the first cut on lathe 1. This is an example of confounding; that is, the effect of *lathe* and *learning* cannot be separated. Hence there is a biased estimate of the effect of the lathe.

To improve the design of this simple experiment, many people would completely randomize the order of cutting the patterns. One possible layout of the completely randomized designed experiment is as follows:

	Pattern															
Lathe	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	27	9	7	15	24	19	10	14	13	8	29	12	21	11	28	17
2	3	25	2	16	22	1	20	31	23	5	26	4	30	18	32	6

The numbers inside the table indicate the order in which the patterns are to be cut on the particular lathe. For example, the first cutting would be pattern number 6 on lathe 2, and the last one (32) would be pattern number 15 on lathe 2.

Although complete randomization provides a more thoroughly designed experiment, one can easily see that peculiar sequences can be obtained. Notice in this so-called completely randomized design that lathe 2 is used to cut first for the first six times and that lathe 1 is used first for the next nine times. It is not known whether this sequencing interferes with the correct decision to buy the better lathe or not, but it is known (mathematically) that complete randomization does provide unbiased estimates of the effects.

Another way to run this experiment (the best way, we think) is to use the following layout:

	Sequence	
	1 Patterns	2 Patterns
Cut	4 7 15 6 3 14 1 10	16 2 9 5 12 13 8 11
First	Lathe 2	Lathe 1
Second	Lathe 1	Lathe 2

The operational procedure for this approach is more complicated, and the need for the additional detail is hard to explain to some experimenters. It is necessary to run the same number of patterns first on lathe 1 as on lathe 2 in order to obtain an unbiased estimate of the difference in time required for the two lathes. We insist on randomly assigning the various patterns to the two different sequences so that one lathe will not be favored over the other as the result of unsuspected differences among the patterns. One possible selection is the use of patterns 4, 7, 15, 6, 3, 14, 1, and 10 for sequence 1, as is shown in the preceding layout. One additional precaution must be taken in order to guard against such effects as fatigue. This can be accomplished by randomly selecting the order in which the patterns are actually cut. This can be done by randomly drawing the numbers 1 through 16. One such sequence would be 14, 11, 6, . . . , 10. Thus pattern 14 would be cut on lathe 2 and then on lathe 1; this would be followed by pattern 11 first on lathe 1 and then on lathe 2; and so on (not necessarily alternating).

This last design is discussed in detail later and is called a "crossover" design.

A Large Manufacturing Company

Another example of a designed experiment (this one "illdefined") occurred a number of years ago in a large manufacturing company. A man working in the production area set up an experiment to test a new alloy, possibly one to replace an old one in production. He ran only *one* heat (batch) of metal with the new alloy and another heat with the old one. Taking one ingot from each heat and 30 pieces of metal from each ingot, he proceeded to test each of the 60 pieces for the property in which he was interested. With the data he made a one-way analysis of variance on the alloys, using the pieces within ingots, with 58 degrees of freedom as the error. The results showed that the new alloy was "better" than the old one, and the experimenter convinced the vice president in charge of production to change the production procedures so that the new alloy would be used in the future. Since the experimenter had used a "designed experiment" and had tested the data "statistically," the vice president concluded that there could be no doubt that the new one was better.

The change cost the company $200,000, and after 2 years in the field, there was as much trouble with the product made from the new alloy as there had been with the old product. The vice president was disgusted and called one of the authors of this chapter to say he would never allow his company to use designed experiments again. After some discussion, the vice president allowed the author to talk with the experimenter to find out how the experiment had been conducted.

In wanting to keep the cost of the experiment low, the experimenter had not considered the possibility that the property in which he was interested varied considerably both from heat to heat and from ingot to ingot within a heat. From an analysis-of-variance point of view, his expected mean squares should have been as follows:

Analysis of Variance of Alloy Problem

Source	*df*	EMS
Alloys (A)	1	$\sigma_p^2 + 30\sigma_I^2 + 30\sigma_H^2 + 30\phi(A)$
Pieces (p) in ingots	58	σ_p^2
Total	59	

Hence, rather than testing that the alloy effect was zero $[\phi(A) = 0]$, he was really testing that the total effect for ingots (I), heats (H), and alloy was zero $[30\sigma_I^2 + 30\sigma_H^2 + 30\phi(A) = 0]$. Since the long-run production of the new alloy did not produce the improvement seen in the experiment, $\phi(A)$ must equal zero. Thus it must have been that σ_I^2 or σ_H^2 or both were not zero. The notation utilized here is that ϕ represents a fixed effect, that is, the effect of two specific alloys, and σ^2 represents a random effect, for example, the ingots are a random sample from all possible ingots.

When the results were explained, the vice president was willing to use designed experiments again on this type of problem, but insisted upon having more than one heat for each alloy.

A Manufacturer of Synthetic Hats

Still another example of a designed experiment (an excellent one) occurred in a company fabricating men's synthetic felt hats. The manufacturer had experienced extreme difficulty in producing these hats so that the flocking appeared on the molded rubber base in a uniform fashion in order to simulate the real felt hats. In approaching this problem, a committee was formed, consisting of a development engineer, a manufacturing foreman, a chief operator, a sales representative, and a statistician. The statistician's job was to obtain from the others all possible causes of imperfect hats. Factors that were thrown out for discussion were as follows: thickness of foam rubber base, pressure of molding, time of molding, viscosity of the latex used to glue the flocking to the molded rubber base, age of the latex, nozzle sizes of several different spray guns, direction of spraying, condition of the flocking, speed of drying, and effect of location within the drying furnace.

After considerable discussion, the committee decided that the most serious problems were probably connected with the nozzle size and the pressure under which the latex was sprayed. In arriving at these factors, the committee essentially forced a review of the entire production process, which led to a better understanding of the process and to an eventual solution to the problem. In this review the chief operator brought out the standard operating levels of the nozzle size as well as the pressure under which the latex was sprayed. Talk with the chief operator revealed that the latex pressure varied considerably because of the viscosity of latex, and from this information the pressure levels were eventually obtained. Additional inquiry ascertained that the manufacturing area had two different nozzle sizes that had been used interchangeably; consequently, the two sizes became the basis for these factor levels.

The authors believe that one is almost always able to find realistic levels for all major factors through committee action of this type. Occasionally, considerable effort is needed to find out just how shoddy one's manufacturing operation really is, and this example shows that actual production operations will allow a process to operate at various levels as long as it works. The determination of the optimal levels is, of course, the desired end of the experimental investigation.

As one can imagine in the manufacturing of synthetic hats, the measuring of product quality is a very difficult task. Consequently, the method used in the experiment just described was to grade the finished hat visually on the following items: the hungry appearance of the flocking, the starchy appearance of the flocking, and the appearance of the brim. During the course of the investigation, these responses were found to be essentially independent of each other and consequently could be treated as three separate dependent variables which could be investigated one at a time. The standards for grading each of these variables were arrived at again through committee action, which eventually resulted in a visual display board and gave the inspectors a realistic means of grading each variable.

One of the most difficult problems in certain types of industrial experiments is the specification of a dependent variable. It is usually quite obvious what the variable should be; however, the methods of measurement are sometimes quite difficult. In ideal cases the value would simply be measured by some simple inspection tool, whereas in other cases the value might be almost impossible to measure and would have to be graded by one or more inspectors.

Having agreed on the variables to analyze, the committee decided to include six nozzle locations, each with a high and low latex viscosity and each with a high and low air pressure, plus two types of base material as factors and levels in the experiment. This required a 2^{13} factorial experiment where all possible combinations of the 13 factors were to be considered. After some work, a fractional replicated design was run, requiring only 256 of the (2^{13}) 8192 combinations. This design allowed complete information on all main effects and on two factor interactions. Fractional factorials are described in detail in Anderson and McLean.[1] + 1984.

The experiment required certain blocking procedures (described later in this chapter), and the results were excellent. The company was able to pinpoint the difficulty in the spraying mechanism, make the necessary changes, and produce a profitable product in 6 months. An interesting sidelight to this experiment was that the company had been losing about $8/hat when the experiment began, and the committee had been given a deadline of 1 year to solve the problem or the company would discontinue the product. With the results from this one experiment, the company was able to make a profit, and a competing company bought the whole process within the year.

The next section describes the authors' basic philosophy regarding the design of experiments, which has been used successfully since 1974.

13.4.2 BACKGROUND

One approach to designing experiments in the last decade considers three essential ingredients of a well-designed experiment, which are expressed by Anderson and McLean.[1] These three ingredients, in order of importance, are (1) inference space, (2) randomization, and (3) replication.

Inference Space

"Inference space" is a term that replaces the term usually used by statisticians, "population." It has been the authors' experience that "inference space" demands more attention from the researcher. Inference space means the limits to which the investigator may use the results of an experiment. Common practice at present is for researchers to indicate, before the experiment is set up, how extensively they wish the results to apply. This requires that they define the experimental units that they are to use in the research and that will be the basis for the inferences. They must also define the time interval and the geographical area to which they wish the results to apply and must then decide which levels of all factors they want controlled in the experiment. Ordinarily, more time should be spent on this phase of designing the experiment than on either of the other two phases (randomization and replication), because without the inference space clearly defined, the so-called best-designed experiment may be worthless to the investigator. We define this ingredient (inference space) as a part of the designed experiment. Hence there can be no "best-designed experiment" without a carefully defined inference space.

Randomization

Randomization is the next most important ingredient in designing experiments. It must be present in the experiment for probability statements to be made. Fisher[3] expressed the idea that it is the physical basis of the validity of the test. It is also the basis of the validity of confidence intervals.

Included in the ingredient of randomization is another concept, "restriction" on randomization. To help the reader understand this concept, let us first explain "completely randomized," which means having no restriction on randomization with respect to the source of the experimental unit. This concept can be seen by considering a factor, t, with five levels and three experimental units treated with each of the five levels of factor t completely at random.

Assume there are 15 randomly drawn experimental units from the inference space to be used for the entire experiment. One way to obtain a completely randomized design is to select a random number between 1 and 5, say 2. Then the first experimental unit must receive treatment 2, or the second level of factor t. Select another number between 1 and 5, say 5; then the second experimental unit must receive treatment 5. Continue sampling or drawing random numbers in this manner until the 15 experimental units have all been "treated." This sampling procedure requires that each level of factor t be represented three times in the experiment; the design of this experiment is completely randomized.

The mathematical model for analyzing the data from such an experiment is

$$y_{ij} = \mu + T_i + \epsilon_{(i)j} \qquad i = 1, 2, \ldots, 5; \quad j = 1, 2, 3 \tag{1}$$

where y_{ij} = the response from experimental unit j treated with level i of factor t
μ = overall mean
T_i = effect of the ith level of factor t
$\epsilon_{(i)j}$ = the experimental error caused by the jth experimental unit nested in the ith level of factor t

The assumptions for the analysis of the data in a model such as this are:

1. y_{ij} is a random variable.
2. The variances of the responses within levels of factor t are equal.
3. The model is additive.
4. The experimental error is normal and independently distributed, with mean zero and variance σ^2, or NID(0, σ^2).

This complete randomization assures the experimenter that the experimental units from the inference space have three opportunities for selection for each treatment. These three units for each treatment then allow a measure of the variation within the treatments, which is the representation of the variation across the entire inference space. It follows, then, that the test of significance on treatment effects (T_i in equation 1) must be based on the excessive amount of variation among the means of the treatments over the amount of variation obtained from within the treatments, where the variation within the treatments is accounted for by the variance due to $\epsilon_{(i)j}$ in equation 1. Hence in equation 1 the $\epsilon_{(i)j}$ is the correct error for evaluating T_i.

If, however, the sampling procedure were such that the first random draw (level 2) were used on the first three experimental units, the second draw (level 5) on the next three experimental units, and so on, we would have a restriction on randomization because the randomization procedures were allowed only 5 times, not 15 as is required for complete randomization. If, as is frequently the case, there tends to be similarity between adjacent units in space or time, the variation within the group of three is smaller than the variation between the groups of three. In this design, then, the variation due to treatments is not separable from the variation between groups of three units (treatments confounded with groups). If, then, there is a source of error variation between groups, it will not be possible to test for treatments, as is depicted in the model

$$y_{ij} = \mu + T_i + \delta_{(i)} + \epsilon'_{(i)j} \tag{2}$$

where $\delta_{(i)}$ is the "restriction" error, or that random component due to the ith group of units (note how the subscript is identical to the subscript to T_i, indicating complete confounding) and $\epsilon'_{(i)j}$ is NID(0, $\sigma^2_{\epsilon'}$), the error due the jth unit in the ith group. It is the variation due to $\delta_{(i)}$ that is representative of the variation across the inference space, whereas the variation due to the $\epsilon'_{(i)j}$ in equation 2 represents only a small portion across the inference space. Hence one needs an estimate of σ^2_δ to test for the effect of treatments. Of course $\delta_{(i)}$ has no degree of freedom, which indicates that this is a poor design and should not be used.

Using the algorithm for deriving the EMS described in Chapter 2 of Anderson and McLean,[1] we can show that the correct error term for testing treatments is $\delta_{(i)}$. However, there is no estimate of this error in equation 2 unless the whole experiment is repeated. Hence the restriction on the randomization has caused $\delta_{(i)}$, which in turn tells the experimenter that this sampling procedure is not a good one even before the first observation is made. This, then, allows the experimenter to change the design early, before taking data that will not give a good analysis.

Before explaining the third ingredient of designed experiments, namely, replication, a few words should be said about special cases where randomization may not be required. If an investigator can show by actual experimentation that the results are the same whether randomization takes place or not, it may not be necessary to randomize. This will happen on extremely well controlled experiments only. We are familiar with an example of a laboratory experiment on one cylinder engine gasoline consumption in which it did not matter whether speeds were randomized or taken in order. The reason for this was that the controls on speed were so precise and the setups so repeatable that the errors were identical within the capabilities of the recording equipment. In terms of the model for analyzing the data from this experiment (similar to equation 2), $\delta_{(i)}$ is approximately zero.

With this nonrandom possibility in mind, and knowing an experiment cannot be designed well without knowing the inference space, we rank inference space above randomization in importance when considering ingredients of a well-designed experiment.

Replication

The third ingredient, replication, is quite often required for an estimate of an error term or to provide the basis for making decisions on the importance of factors contributing to the response variables. In addition, as the number of observations increases on a given treatment, the more precise the estimate of the effect of the treatment becomes, or the smaller its variance becomes.

If, however, previous experimentation has shown that certain information is available, for example, that the variance is known or that higher-order terms in the model are zero, it may not be necessary to replicate the entire experiment. In fact, there are many good experiments run with fewer than the total number of levels of the factors in a "factorial" experiment. These experiments are called "fractional replicated factorials."

Because of the various well-designed experiments without complete replications, the ingredient replication is placed in third position behind inference space and randomization.

Readers of the material on designs and analyses in the next section should understand (1) basic statistical concepts, (2) distributions such as mean $\overline{y}$, t, χ^2, and F, and (3) analysis of variance and regression models.

13.4.3 DESIGNS

A definition of a designed experiment is an arrangement of the experimental material, including randomization of experimental units to the treatments, so that statistical tests of significance (and confidence intervals) on the effects and interactions of the factors being studied can be made. To accomplish this, care must be taken to set up the arrangement efficiently (keep the cost reasonably low) while covering the inference space. In the following coverage of designs, major headings are used to indicate designs encountered by the authors in engineering studies.

Block Designs

Importance of Blocking

Many authors of books on the design of experiments express the importance of "blocking," or placing all treatments or all combinations of the levels of all factors of interest in a homogeneous group (thereby removing some of the effect of an extraneous variable from the experimental error) and repeating this group or block in time and/or space with different experimental units. To show this concept, we use mathematical equations that are to be used as the basis for analyses of the data from the designed experiment.

Returning to the concepts involved in setting up equations 1 and 2, it follows that another design of the experiment is to arrange three blocks of five treatments each, where the five experimental units are randomized onto the five treatments per block. It follows that the equation to be used as the basis for the analysis is

$$y_{ij} = \mu + B_i + \delta_{(i)} + T_j + \epsilon_{ij} \qquad (3)$$

where y_{ij} and μ have the same meaning as they did in equation 2, $\delta_{(i)}$ is similar to $\delta_{(i)}$ in equation 2, and

B_i = effect of block i
T_j = effect of the jth treatment
ϵ_{ij} = error due to the jth treatment in block i (assuming there is no interaction)

Since the experimenter is interested in testing for treatment effects only, this is an excellent design because ϵ_{ij} is the basis for the test of T_i since the variation across treatments is compared to the variation due to ϵ_{ij}'s. If it should turn out that the effects of B_i and $\delta_{(i)}$ are zero, then B_i, $\delta_{(i)}$, and ϵ_{ij} may be pooled, and equation 3 becomes equation 1. This completes the demonstration that, in general, blocking is always worthwhile in experiments and should be used whenever possible.

Latin square designs, used correctly, merely extend restriction on randomization in one more dimension. Too many engineers *incorrectly* use Latin square designs as special fractional factorials (see Chapter 8 in Anderson and McLean[1]). Since it is difficult to enable the reader to understand these points in the space allowed here, we recommend that Latin squares not be used by novices.

Incorrect Use of Blocks

In the previous example, the effect of $\delta_{(i)}$, the restriction error caused by blocks, did not decrease the importance of blocking because the experimenter was interested in the effect of blocking plus the restriction error in reducing the estimated experimental error, $\hat{\epsilon}_{ij}$. In that case $\hat{\epsilon}_{ij}$ was the basis for testing treatment effects only.

One must consider the case in which the blocking concept is used incorrectly. Consider an example in evaluating microforms, where the interest was in reducing user dissatisfaction. The experimenter wanted to use two types of projection (front, F, and back, B) as blocks and to randomize four screen angles (0°, 45°, 90°, 105°) twice within each block. The measured variable was seconds required to read the material presented each time. Pictorially the design was as follows:

Types of Projection	
Front	Back
90°	45°
0°	105°
45°	0°
90°	105°
105°	90°
0°	45°
45°	0°
105°	90°

The correct degrees of freedom and the model for the analysis of the data from this design are

$$16 = 1 + 1 + 0 + 3 + 3 + 8$$

$$y_{ijk} = \mu + T_i + \delta_{(i)} + A_j + TA_{ij} + \epsilon_{(ij)k} \qquad i = 1, 2; \quad j = 1, 2, 3, 4; \quad k = 1, 2 \tag{4}$$

where y_{ijk} = seconds required to read the microform from the kth observation using the jth angle and the ith type of projection

μ = overall mean

T_i = effect of the ith type of projection (fixed)

$\delta_{(i)}$ = restriction error due to all of the angles used with the ith type of projection before the other type of projection is used

A_j = effect of the jth angle (fixed)

TA_{ij} = effect of the interaction of the ith type with the jth angle

$\epsilon_{(ij)k}$ = error due to the kth observation within the ith type jth angle, assumed NID(0, σ^2).

The analysis of variance appropriate for analyzing data using equation 4 is as follows:

Source	*df*	EMS
Types of projection (T_i)	1	$\sigma_\epsilon^2 + 8\sigma_\delta^2 + 8\phi(T)$
Screen angles (A_j)	3	$\sigma_\epsilon^2 + 4\phi(A)$
TA_{ij}	3	$\sigma_\epsilon^2 + 2\phi(TA)$
Repeats within ($T - A$) Combination $\epsilon_{(ij)k}$	8	σ_ϵ^2

It is apparent from the EMS column that there is no test for types of projection because there is no source with an EMS of $\sigma_\epsilon^2 + 8\sigma_\delta^2$. Hence this is an incorrect use of blocks. To obtain the source to estimate $\sigma_\epsilon^2 + 8\sigma_\delta^2$, there must be a replicate of the experiment. This concept is demonstrated in the following section.

Correct Use of Blocks

To make the experiment described in the preceding section a good one, the investigator should set up the eight treatment combinations as follows (one possible randomization):

Blocks			
1		2	
F	45°	B	105°
B	0°	F	90°
B	105°	B	45°
F	0°	F	105°
B	90°	F	45°
F	105°	B	0°
B	45°	B	90°
F	90°	F	0°

The degrees of freedom and the model for the analysis of the data from this design are as follows:

$$30 = 1 + 1 + 0 + 1 + 1 + 3 + 3 + 3 + 3 + 0$$

$$y_{ijk} = \mu + B_i + \delta_{(i)} + T_j + BT_{ij} + A_k + BA_{ik} + TA_{jk} + BTA_{ijk} + \epsilon_{(ijk)}$$

$$i = 1, 2; \quad j = 1, 2; \quad k = 1, 2, 3, 4$$

where

B_i = effect of ith block (random)
$\delta_{(i)}$ = restriction error using blocks correctly
T_j, A_k, TA_{jk} = same as T_i, A_j, TA_{ij} of equation 4
$BT_{ij}, BA_{ik}, BTA_{ijk}$ = appropriate errors for testing (note arrows), which would probably pool for a common error
$\epsilon_{(ijk)}$ = residual error, not estimable with one observation per treatment combination per block

This is explained in detail in Anderson and McLean.[1]

Repeated Measures and Crossover Designs

A researcher in an industrial plant dealing with soldering parts on electronic equipment was interested in studying paced and unpaced production. He set up the experiment in a plant that used women only and recorded the ages of the women. He grouped the women into ages as follows: young (18–23), middle (30–35), and old (52–57). He was able to obtain six women in each of these age groups, which allowed him to have a sequence (selected at random) of (paced, unpaced) and (unpaced, paced) for 2 weeks for three women in each group. The layout of the design of the experiment can be portrayed as follows:

	Ages					
	Young		Middle		Old	
	Sequence		Sequence		Sequence	
	1	2	1	2	1	2
	Women	Women	Women	Women	Women	Women
Weeks	1 2 3	4 5 6	7 8 9	10 11 12	13 14 15	16 17 18
1	P	U	U	P	U	P
2	U	P	P	U	P	U

If the assumptions of the analysis of variance have been met, the analysis of variance of the data (different response variables) from this experiment, which uses a crossover design, is as follows:

Source	df
Ages (A)	2
Sequence (S)	1
AS	2
Females in (A – S) cells (F)	12
Weeks (W)	1
Paced versus unpaced (P)	1
AP	2
SP	1
ASP	2
Residual	11

Refer to Grizzle,[4] Myers,[5] and Anderson and McLean[1] for details of this type of design and analysis. For designs and analyses utilizing extensions of these designs to more sequences, refer to Albert et al.[6] and Westlake.[7]

Other Designs

There are many other designs used in industrial engineering that deal with factorial experiments and fractional factorials plus response surface designs. For these designs and analyses, refer to Myers,[8] Anderson and McLean,[1] and Box et al.[9]

REFERENCES

1. V. L. ANDERSON and R. A. MCLEAN, *Design of Experiments: A Realistic Approach*, Dekker, New York, 1974.
2. W. G. COCHRAN and G. M. COX, *Experimental Designs*, 2nd ed., Wiley, New York, 1957.
3. R. A. FISHER, *The Design of Experiments*, 7th ed., Hafner, New York, 1960.
4. J. E. GRIZZLE, "The Two-Period Change-Over Design and Its Use in Clinical Trials," *Biometrics*, Vol. 21, 1965, pp. 467–480.
5. J. L. MYERS, *Fundamentals of Experimental Design*, 3rd ed., Allyn & Bacon, Boston, 1979.
6. K. S. ALBERT, S. W. BROWN, JR., K. A. DESANTO, A. R. DISANTO, R. D. STEWART, and T. T. CHEN, "Double Latin Square Study to Determine Variability and Relative Bioavailability of Methylprednisolon," *Journal of Pharmaceutical Sciences*, Vol. 68, No. 10 (October 1979), pp. 1312–1316.
7. W. J. WESTLAKE, "Statistical Aspects of Comparative Bioavailability Trials," *Biometrics*, Vol. 35, 1979, pp. 273–280.
8. R. H. MYERS, *Response Surface Methodology*, Allyn & Bacon, Boston, 1971.
9. G. E. P. BOX, W. G. HUNTER, and J. S. HUNTER, *Statistics for Experimenters*, Wiley, New York, 1978.

[redacted]

10, WINER, B. J., *Statistical Principles in Experimental Design*, 2nd ed., McGraw-Hill, New York, 1971.

11. McLean, R.A. and Anderson, V.L., "APPLIED FACTORIAL and FRACTIONAL DESIGNS", Marcel Dekker, 1984.