

BASIC EXPERIMENTAL DESIGNS OF INDUSTRIAL EXPERIMENTS

BY

DR. MARCELO M. ORENSE

ASST. PROFESSOR OF STATISTICS

U. P. STATISTICAL CENTER

1. Introduction

[The following paper is essentially that prepared and read at a seminar of the Philippine Chemical Society. However, some numerical examples of the analysis are added which an industrial technologist may find convenient and helpful for the interpretation of results.]

Several research projects require the collection of data. In chemistry and industry this is usually done by conducting experiments which however should be designed to provide maximum information under certain restrictions of fixed amount of time, effort and money and should also lead to valid conclusions.

This paper is intended to introduce the basic concepts of experiments involving several factors. However, any such introductory discussion must also include the basic experimental designs. The present treatment of this subject matter is by no means complete and it is mostly descriptive in nature to interest chemists and other technologists with limited statistical background to look into the possibility of giving statistics an opportunity to help make their work more interesting or enlightening if not more profitable in some aspects.

Designing an experiment involves the principles of replication, randomization and the design proper. Replication denotes the execution of an experiment more than once so as to increase precision and to supply an estimate of the experiment-

al error. It is replication which makes a test of significance possible. Valid conclusions are generally made after a test of significance is performed.

The word **treatment** will be used to apply to temperature, pressure, chemical fertilizers or machines used in an experiment. A set of treatments applied to a set of experimental units is said to be **randomized** when the treatment applied to a particular unit is chosen at random by use of draw lots or table of random numbers from the set of treatments. Any two treatments are afforded by randomization an equal chance to be assigned to adjacent or non-adjacent units in the design, hence, any correlation between any two treatments tends to cancel out. Randomization has been likened by Cochran and Cox (1957) to "insurance against disturbances that may or may not occur and that may or may not be serious if they do occur."

2. Basic Experimental Designs

With the advent of modern statistics, both applied and mathematical, many experimental designs are available of which the most commonly used are briefly mentioned here. The existence of more complicated designs for more complex situations will also be indicated briefly. The choice of a design for a particular experiment depends largely on the restrictions which may be placed on the assignment of the treatments to the experimental units. Also, the choice involves the ability and familiarity of an experimenter to recognize all the possible components contributing to the variability of the response to be measured and thereby making the design in such a manner that some undesirable components should be eliminated if possible. If not possible, then the true effects of all components affecting the measured response in a properly

1 A test of significance is one which, by use of a test criterion, provides a test of hypothesis that an effect is zero.

designed experiment could be estimated by known statistical methods.

The simplest type of experimental design is known as the **completely randomized design**. It is applicable when the whole set of experimental units (or materials) are homogeneous. For example, let an experiment be conducted to evaluate the performance of three proposed additives to gasoline to improve the performance of cars, say, in terms of mileage. Assume that it is possible to make measurements to this effect. Suppose we have nine cars of same make and size. Let the treatments (additives) be T_1, T_2, T_3 , where one of these may be a control. These are to be assigned to the nine cars at random whereby three cars are to receive the same treatment. In this setup, we assume the cars to be more or less alike. It becomes necessary also to assume that nine drivers selected have more or less the same driving ability. In a completely randomized experiment, however, there need not be equal number of experimental units (cars) for the treatments. This unequal replication would allow some treatment effects having more replication to be estimated more precisely than others with less replication. It is to be pointed out however, that equal replication provides a more powerful test.² The observed measurements (Y) on this simple experiment can be classified in one way, i.e., according to the treatment used:

T_1	T_2	T_3
-	-	-
Y_{11}	Y_{21}	Y_{31}
Y_{12}	Y_{22}	Y_{32}
Y_{13}	Y_{23}	Y_{33}

² A test is said to be more powerful than another if the probability that it rejects the alternative hypothesis when false is greater than that of the other.

With these data one can proceed to estimate the mean value of the variable Y , the effects of the treatments and to test if significant differences exist among the three additives. A researcher could then choose the best additive as far as car performance is concerned. Usually, a choice has to depend also on the cost involved.

The second type of design which is the most popularly used is the **randomized complete blocks**. When the experimental units exhibit a known pattern of heterogeneity such that they can be classified into equal groups (or better known as "blocks" in statistics), each consisting of homogeneous units, then this design can be used provided that the number of units in each block is equal to the number of treatments to be used. It is seen therefore that this design requires an additional restriction to that of the completely randomized. With the three treatments T_1, T_2, T_3 , suppose there are nine cars divided equally into three categories, (maybe same "make" but different sizes or same size but different "makes"). In the randomized complete blocks design the treatments are allotted at random to the three cars in each category. Data resulting from such a design can be arranged in a two-way classification, namely according to block and treatment:

		Treatment (Additive)		
		T_1	T_2	T_3
Block (Car category)	I	Y_{11}	Y_{21}	Y_{31}
	II	Y_{12}	Y_{22}	Y_{32}
	III	Y_{13}	Y_{23}	Y_{33}

The statistical analysis of these data can separate and estimate the effects of treatments and blocks. Differences among the three treatments would be free from effects attributed to the

car categories. A test to determine significance of such differences among treatments becomes meaningful. Valid conclusions regarding these differences can therefore be made.

The third type called **Latin Square**, exhibits more restrictions on the plan. The pattern of heterogeneity in the experiment may be such that two ways of blocking are possible, say into equal number of rows and columns. Furthermore, the number of units or trials in a row or column should be equal to the number of treatments. For example, with the same three treatments, assume that the trials will be made on 3 cars of same "make" and size. Let us say that three drivers can run three trials per day, i.e., morning, afternoon, evening.

		Drivers		
		1	2	3
Time	Morning	T ₁	T ₂	T ₃
	Afternoon	T ₂	T ₃	T ₁
	Evening	T ₃	T ₁	T ₂

Standard Plan of a 3 x 3 Latin Square

It is observed in the plan that a treatment is used once and only once in a row (time) and once in a column (driver). This is a characteristic feature of the Latin Square design. Data from such a design can be classified according to treatment, row, or column. The analysis of such data includes estimating an effect of treatment, time of trial and driver and also test of significance of differences among treatments. Tests on differences among mean responses for different times of trial or drivers may also be done if desired. The additional features, like time of trial and drivers whose effects may not be necessary but whose presence in the design may effect the measured response, could be separated from the true effects of the treatments by this careful planning. Fisher and Yates

(1957) give a list of Latin Squares and their randomization procedures.

3. Analysis of Variance

The analysis of variance for each of these designs can easily be learned from standard textbooks in statistical analysis. This analysis is an arithmetic procedure of partitioning the observed variation of the data into the individual contributions of the components involved in the experiment. What is given here outlines the steps involved when the design used is the randomized complete blocks. The analysis for the other designs can be done in a similar way.

In the randomized complete blocks design an observation in the j th block receiving the i th treatment, to be denoted by Y_{ij} , is usually represented as the sum of four terms: the general mean m , an effect t_i due to the i th treatment, an effect b_j due to the j th block, and a random error e_{ij} which is a residual and may consist of all remaining components not explained by the first three terms. This relationship is ordinarily given as a linear additive model.

$$Y_{ij} = m + t_i + b_j + e_{ij},$$

$i = 1, 2, 3, \dots, p$ treatments

$j = 1, 2, 3, \dots, r$ blocks.

The estimates m, t_i and b_j are obtained by a least squares procedure which minimizes the sum of squares due to error.³ These estimates are computed as follows:

$$m = G/rp, \text{ where } G \text{ is the total of all observations,}$$

$$t_i = \frac{T_i}{r} - m, \text{ where } T_i \text{ is total for } i\text{th treatment,}$$

$$\text{and } b_j = \frac{B_j}{p} - m, \text{ where } B_j \text{ is total for } j\text{th block.}$$

³ Estimates obtained by this procedure are "best" linear estimates and whose values are such that the sum of squares due to error is a minimum.

The formulas for the treatment and block effects show these as differential effects between the treatment or block means and the mean yield m .

The test of a null hypothesis that the treatment effects are zero (i.e., treatment means are equal) is included in an analysis of variance. This analysis is tabulated as follows:

**ANALYSIS OF VARIANCE
FOR A RANDOMIZED COMPLETE BLOCKS DESIGN**

Sources of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F
Blocks	$r - 1$	SS(B)	MS(B)	
Treatments	$p - 1$	SS(T)	MS(T)	MS(T)/MS(E)
Error	$(r - 1)(p - 1)$	SS(E)	MS(E)	

Total	$rp - 1$	SS(Total)
-------	----------	-----------

Where $SS(B) = \sum_{j=1}^r \frac{B_j^2}{p} - \frac{G^2}{rp}$. sum of squares due to blocks,

$SS(T) = \sum_{i=1}^p \frac{T_i^2}{r} - \frac{G^2}{rp}$. sum of squares due to treatments,

$SS(Total) = \sum_{j=1}^r \sum_{i=1}^p y_{ij}^2 - \frac{G^2}{rp}$. sum of squares due to all observations,

and $SS(Error) = SS(Total) - SS(B) - SS(T)$.

It can be seen that the total sum of squares, $SS(Total)$, has been partitioned into three sums of squares corresponding to the three sources of variation of the measured responses.

The entries in the column for mean squares are obtained by dividing the sums of squares by their respective degree of freedom.

The computed value of F given by $MS(T)/MS(E)$ provides a means of determining whether the differences among the treatment means are significant or not at a pre-determined level of significance. This observed value of F is compared with its tabular value (value of F if the null hypothesis is true) at the given level of significance. An observed value greater than the tabular indicates significant differences among treatment means.

An analysis of variance for the completely randomized design will have as sources of variation the treatments and error while for a Latin Square the rows, columns, treatments and error.

4. Numerical Example

Let the analysis be illustrated by the following data which are expressed in kms./liter. These are observed on nine cars belonging to three categories (blocks) used in an experiment to estimate the effects of three additives (treatments) on gasoline. One of these treatments may be a "control, i.e., without an additive.

		TREATMENTS (GASOLINE ADDITIVES)				
		T_1	T_2	T_3	Total	Mean
Blocks (Categories)	B_1	16.6	10.6	9.5	36.7	12.23
	B_2	15.3	10.2	8.8	34.3	11.43
	B_3	10.9	6.0	4.2	21.1	7.03
of car	Total	42.8	26.8	22.5	92.1	
	Mean	14.27	8.93	7.50	10.23 = m	

Let it be assumed that an observed response which depends on a mean response m , a differential effect of the additive used (t_i), another effect due to the category of the car (b_j),

and a residual effect called error (e_{ij}), be represented by the following model:

$$Y_{ij} = m + t_i + b_j + e_{ij}$$

It is implied here that the effects t_i , b_j and the residual e_{ij} are additive, e.g., a treatment simply adds up an effect to the mean response. The residual e_{ij} , the part of the observed response which is "unexplained", may partly be due to the assumption of such simple model and partly due to random fluctuations which can never be avoided.

The estimates of the effects of treatment and car categories are given as follows:

<u>Treatment Effects</u>	<u>Effects of Car Categories</u>
$t_1 = + 4.04$	$b_1 = + 2.00$
$t_2 = - 1.30$	$b_2 = + 1.20$
$t_3 = - 2.74$	$b_3 = - 3.20$

When these effects are combined with the mean response, in an additive manner, i.e., $m + t_i + b_j$, the sum is the "explained" part of the observation. The "unexplained" part is the difference of the actual observation and the "explained" portion.

	<u>Explained ($m + t_i + b_j$)</u>			<u>Residual ($Y_{ij} - m - t_i - b_j$)</u>		
	T_1	T_2	T_3	T_1	T_2	T_3
B_1	16.27	10.93	9.49	+ 0.33	- 0.33	.01
B_2	15.47	10.13	8.69	- 0.17	0.07	.11
B_3	11.07	5.73	4.29	- 0.17	0.27	-.09

The rows and columns for residual should add up to zero. In this example, they do not add up to zero because of rounding errors.

The analysis of variance indicated below shows that there exists evidence of significant differences among the means of the treatments and also among the categories of cars. This can also be stated in terms of the effects, i.e., the effects due to treatments and car categories are significantly different from zero.

ANALYSIS OF VARIANCE

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	
Treatments	2	76.29	38.14	Significant
Categories	2	47.04	23.52	Significant
Error	4	0.37	0.0925	
Total	8	123.70		

The estimate of the standard error of a treatment mean or effect is equal to 0.17.

$$\sqrt{\frac{.0925}{3}} = 0.17.$$

From the estimates of the treatment effects an industrial technician may realize the technological importance of these effects.

5. Factorial Experiments

In the preceding discussion the treatments belonged to one factor, i.e., additive to a gasoline. A factor is defined as a kind of treatment. A factor consists of different levels, either quantitative or qualitative values. For example, the factor additive may have two levels consisting of two different amounts of a particular additive, or two different types of additives.

In many experiments two or more factors usually enter into consideration. A treatment or treatment combination is

then made up of the levels of each of the factors. Let us say factor A, one kind of the additive to the gasoline, consists of two amounts to be denoted by a_1 and a_2 . Let another factor be a second type of additive; call this factor B and consisting also of two levels, b_1 and b_2 .

A complete factorial experiment, i.e., consisting of all the possible combinations of the levels, would have the following four treatment combinations: a_1b_1 , a_1b_2 , a_2b_1 , a_2b_2 . Such a factorial arrangement can then use any appropriate experimental design, e. g., the randomized complete blocks. A block in such a design would have four plots⁴ (cars in the example) to which the four treatment combinations are randomly assigned.

Many of us are familiar with the classical method of varying one at a time and holding all others constant, e.g., using b_1 and varying the A-factor. The main objection to such method arises when interaction between A and B exists, i.e., a level of A may be best for one level of B and another level of A is best for another level of B. Which level of B should then be used with the levels of A in such a classical experiment?

With modern statistical methods it is possible to study the interplay of all the factors simultaneously and determine what combination of levels of these factors provides the optimum response. Factorial experiments can give us not only the differential effect of each factor (this is what the classical experiment could give us) but at the same time the interaction between two or more factors. Interaction is defined as a measure of the extent to which the effect of changing the level of one factor depends on the levels of other factors. It is the estimation of this interaction between two or more factors that renders factorial experiments indispensable. In fact, this type of experiments becomes extremely useful in

⁴ The agricultural connotation should be ignored when trial is other than agricultural.

exploratory experimental trials designed to find the proper combinations of the levels of several factors that would optimize the response.

6. Analysis of $p \times q$ Factorial Experiments

In an experiment with two factors with p and q levels, respectively, we would be able to estimate the effects of the individual amounts of the factor A, the second factor B, interaction between these two factors, and also the blocks used in the design. The main purpose in such an experiment would be to study the interaction. If found insignificant, the individual effects of the factors may still be of some interest. Proper recommendation as to which level of the A-factor to use in conjunction with a level of B could be made after tests of significance. The analysis is based on the model showing the relationship between an observed response and the different effects which affect the measured response.

$$Y_{ijk} = m + r_k + a_i + b_j + (ab)_{ij} + e_{ijk}$$

$i = 1, 2, \dots, p$ amounts of A

$j = 1, 2, \dots, q$ amounts of B

$k = 1, 2, \dots, r$ blocks,

where m = mean response

r_k = effect of k th block

a_i = effect of i th level of factor A

b_j = effect of j th level of factor B

$(ab)_{ij}$ = interaction between i th level of A and j th level of B

e_{ijk} = random error on the (ijk) th observation.

The estimates are obtained as follows:

$$m = G/pqr \text{ where } G = \text{total of all } pqr \text{ observations.}$$

$$a_i = \frac{A_i}{qr} - m, \quad A_i = \text{total for } i\text{th level of factor A}$$

$$r_k = \frac{R_k}{pq} - m, \quad R_k = \text{total for } k\text{th block}$$

$$b_j = \frac{B_j}{pr} - m, \quad B_j = \text{total for } j\text{th level of factor B}$$

$$\text{and } (ab)_{ij} = \frac{(AB)_{ij}}{r} - \frac{A_i}{qr} - \frac{B_j}{pr} + m.$$

where $(AB)_{ij}$ = total of all observations to which the treatment combination $a_i b_j$ is applied.

The hypothesis that the effects of the p levels of A and those of B are zero can be tested in an analysis of variance. Similarly, the hypothesis of no interaction between A and B can also be tested.

Analysis of Variance of $p \times q$ Factorial Experiment in a Randomized Complete Blocks Design

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F
Blocks	$r-1$	$SS(R)$	$MS(R)$	
A	$p-1$	$SS(A)$	$MS(A)$	$MS(A)/MS(E)$
B	$q-1$	$SS(B)$	$MS(B)$	$MS(B)/MS(E)$
AB	$(p-1)(q-1)$	$SS(AB)$	$MS(AB)$	$MS(AB)/MS(E)$
Error	$(pq-1)(r-1)$	$SS(E)$	$MS(E)$	
Total	$pqr-1$	$SS(\text{Total})$		

The sums of squares are computed as in Sec. 3.

$$SS(R) = \sum_k \frac{R_k^2}{pq} - \frac{G^2}{pqr}$$

$$SS(A) = \sum_i \frac{A_i^2}{qr} - \frac{G^2}{pqr}$$

$$SS(B) = \sum_j \frac{B_j^2}{pr} - \frac{G^2}{pqr}$$

$$SS(AB) = \sum_{ij} \frac{(AB)_{ij}^2}{r} - \frac{G^2}{pqr} - SS(A) - SS(B)$$

$$SS(\text{Total}) = \sum_{ijk} Y_{ijk}^2 - \frac{G^2}{pqr}$$

and $SS(E)$ is obtained as a difference of $SS(\text{Total})$ and the other four sums of squares. The computed values of the test-criterion F compared with corresponding tabular values will show whether the effects of the levels of A , of B , and the interaction between A and B , respectively, are significantly different from zero or not at a predetermined level of significance. A greater computed value than the tabular value of F indicates significance, hence, the rejection of the hypothesis.

Numerical Example.

The analysis is being illustrated by a 2×2 factorial experiment. The observations are as follows:

		Treatments				Total
		a_1b_1	a_1b_2	a_2b_1	a_2b_2	
Blocks	I	3.6	7.1	24.2	6.3	41.2
	II	3.5	5.1	16.0	4.5	29.1
Total		7.1	12.2	40.2	10.8	70.3

An observation Y_{ijk} may then be expressed by the model given in Sec. 6.

A table of means of the four treatment combinations is given below:

	α_1	α_2	
b_1	3.5	20.1	11.8
b_2	6.1	5.4	5.8
	4.8	12.8	8.8

From these means the estimates of effects are obtained.

Effects of levels of A:

$$\alpha_1 = - 4.0$$

$$\alpha_2 = + 4.0$$

Effects of levels of B:

$$b_1 = + 3.0$$

$$b_2 = - 3.0$$

Interaction effects:

$$\alpha_1 b_1 = - 4.3$$

$$\alpha_2 b_1 = + 4.3$$

$$\alpha_1 b_2 = + 4.3$$

$$\alpha_2 b_2 = - 4.3$$

The block effects may also be estimated:

$$r_1 = \frac{41.2}{4} - 8.8 = 1.5$$

$$r_2 = \frac{29.1}{4} - 8.8 = - 1.5$$

When these effects are combined as in the assumed model, the combination consists of the "explained" and the "unexplained" or residual portions:

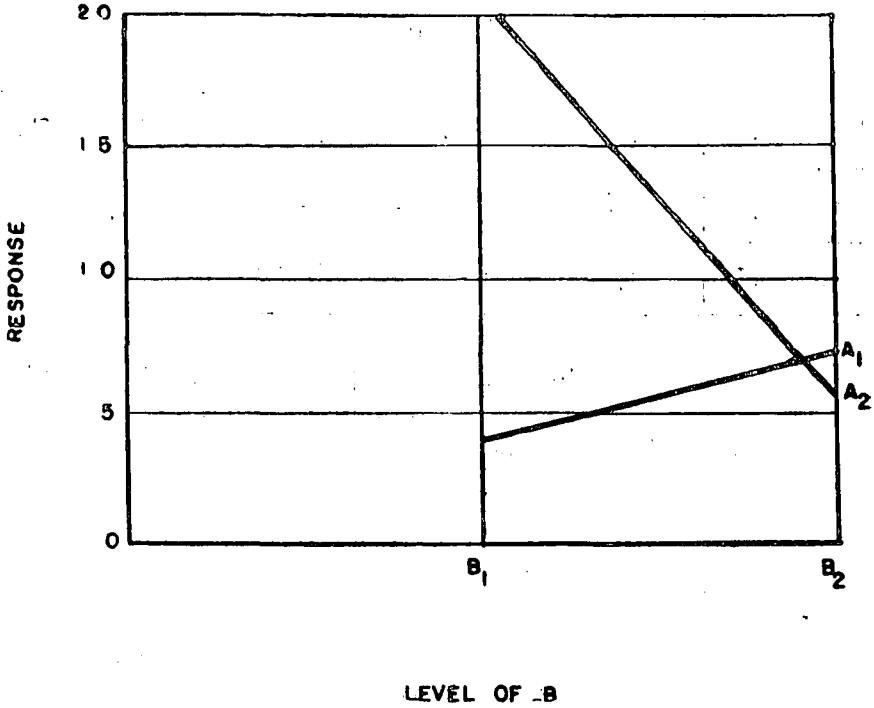
Treatment Combination	Explained $m + r_k + a_j + b_j + (ab)_{ij}$				Unexplained (Residual)			
	a_1b_1	a_1b_2	a_2b_1	a_2b_2	a_1b_1	a_1b_2	a_2b_1	a_2b_2
Block I	5.0	7.6	21.6	7.0	-1.4	-0.5	+2.6	-0.7
Block II	2.0	4.6	18.6	4.0	1.5	0.5	-2.6	0.5

The sum of each column and row for the residuals should add up to zero. There are some rounding errors here involved. The analysis of variance is given as follows:

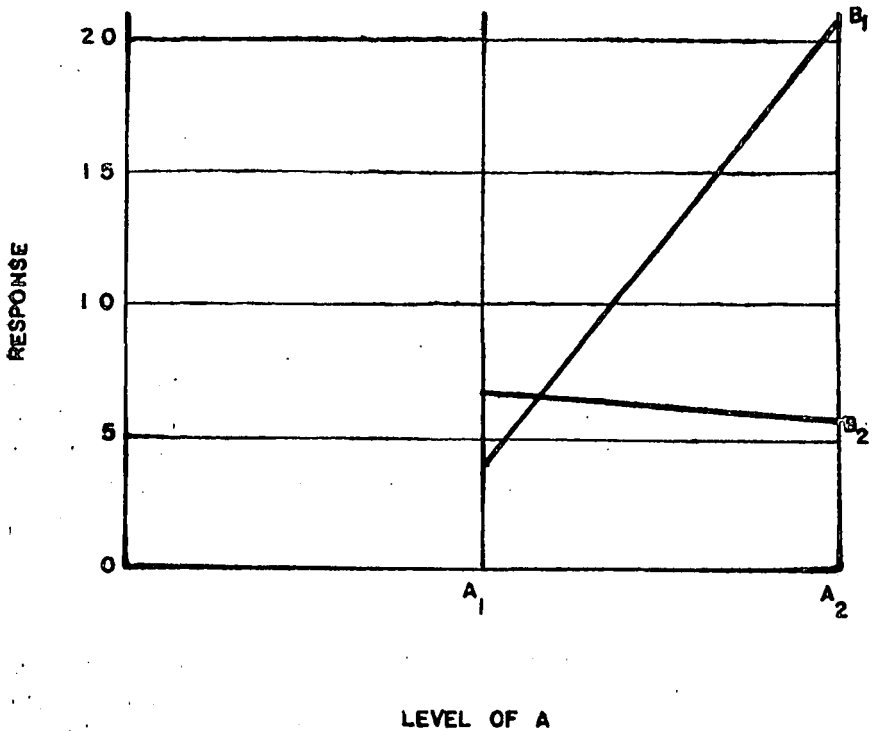
Source of Variation	d.f.	Sum of Squares	Mean Squares	F
Blocks	1	18.3013	18.3013	
A	1	125.6113	125.6113	19.89*
B	1	73.8113	73.8113	11.69*
AB	1	148.7812	148.7812	23.56*
Error	3	18.9437	6.3146 = S^2	
Total	7	385.4488		

The computed F-values for A and B both exceed the tabular values of F at 5% level of significance. This means that there exists enough evidence to show that the treatment effects are not zero.

The interaction between A and B is also significant at 5%. To explain this interaction consider the following graph of the means:



One can see from this graphical representation the effect of interaction, (i.e., the effect of changing the level of one factor as dependent on the other factor). For a fixed level of A, there is a change in the means as B increases. The mean response decreases abruptly for a_1 and increases slightly for a_2 as B increases. This change is both in amount and direction. This interaction is a symmetrical relationship of A and B. The mean response has a rapid rise for b_1 and gradual decrease for b_2 as A increases.



Other Aspects of Experiments

Complete factorials although useful in several cases have some limitations. If the factors involve several levels, the number of treatment combinations becomes very large. If a large variability among the experimental units exists, the grouping of these units into more or less homogeneous blocks may not be practical. Hence, if these blocks are to be made

homogenous it then becomes necessary to include in a block fewer units than the number of treatments. Of course the number of factors may be decreased, but if it is imperative to study all the factors in the same experiment, then one may resort to the methods of confounding or fractional replication

Under certain situations, the size of a block should be reduced to maintain the desired homogeneity of the units within a block. For example, in an experiment consisting of four factors each at two levels there would be 2^4 (or 16) treatments. A complete factorial would require a block of 16 experimental units. By the confounding method, one could use blocks of size 8 which then could have only 8 of the possible 16 treatments, and the other half of the treatments are used in another block of size 8. Confounding implies that a treatment effect (usually chosen to be a certain interaction) is mixed up with the block effects, hence cannot be estimated from the data. The other factorial effects however, could be estimated. A design could be so planned that it consists of more than one replication, where an interaction could be confounded in a replication but estimable in another replication where some other interaction is confounded.

The other possibility is called fractional replication, i.e., it may be desired to use only a fractional of the treatment combinations. For example, a half replicate of 2^3 design or one quarter of 2^5 can be planned for an experiment. Such a method also involves a loss of information due to confounding of many effects but enables the estimation of other necessary effects which are more important than the confounded ones. This differs from the confounding method in that only a fraction of the treatment combinations appears in a design, whereas in confounding, the whole set of treatment combinations are used, but they appear in different blocks. Although some information is lost in these two methods, a designer of an experiment can choose which effect or effects are not so important and hence can be sacrificed, and can still study all the factors in a single experiment.

Other designs belonging to the type of incomplete blocks, either balanced or so-called partially balanced, are available for use although they involve more complicated statistical analysis. One very popularly used design is the so-called split-plot design. These are mentioned here just to convey the information that there are many possible designs that can be used for most types of experimental work.

The presentation of these few descriptive aspects of designs of experiment, I hope, would encourage chemical researchers and other technologists to plan more carefully their experiments in order to obtain the necessary information. Let us remember that gathering data is not the mere act of collecting; the right kind should be collected in the right manner.

LIST OF REFERENCES

1. Anderson, R. L. (1957) Complete factorials, fractional factorials and confounding, Proc. Symposium on Design of Industrial Experiments, Nov. 1956, Institute of Statistics, Raleigh, North Carolina.
2. Anderson, R. L., and Bancroft, T. A. (1952) Statistical Theory in Research, McGraw-Hill Book Co., Inc. New York.
3. Bennett, C. A., and Franklin, N.L. (1954) Statistical Analysis in Chemistry and the Chemical Industry, John Wiley and Sons, Inc., New York.
4. Bose, R. C., Clatworthy, W. H., and Shrikhande, S. S. (1954) Tables of Partially Balanced Designs with Two Associate Classes, North Carolina Agri. Exp. Sta. Tech. Bul. No. 107.
5. Box G. E. P. (1954) The exploration and exploitation of response surfaces: Some general considerations and examples, Biometrics 10:16-60.
6. Box, G. E. P., and Hunder, J. S. (1957) Experimental designs for the exploration and exploitation response surfaces, Proc. Symposium on Design of Industrial Experiments, Nov., 1956, Institute of Statistics, Raleigh, North Carolina.
7. Box, G. E. P., and Wilson, K. P. (1951) On the experimental attainment of optimum conditions, Jour. Roy. Stat. Soc., Series B, 13: 1-45.
8. Brownlee, K. A. (1953) Industrial Experimentation, Chemical Publishing Co., Inc., New York.
9. Cochran, W. G., and Cox, G. M. (1957) Experimental Designs, John Wiley and Sons, Inc., New York.
10. Fisher, R. A. and Yates, F. (1957) Statistical Tables for Biological Agricultural and Medical Research, Hafner Publishing Co., Inc., New York.