



A SunCam online continuing education course

What Every Engineer Should Know About the Design and Analysis of Engineering Experiments II

by

O. Geoffrey Okogbaa, Ph.D. PE



The Design and Analysis of Engineering Experiments II
A SunCam online continuing education course

Contents

Introduction 4

1.1 Role of Design of Experiments in Quality Design and Improvement..... 4

1.2 The Role of Design of Experiments in the Design Process..... 5

Missing Values 5

2.1 Missing values for Single Factor Randomized Block Design (RBD) 6

2.2 A Note about the Adjusted ANOVA Table due to Missing Value..... 9

2.2.2 Missing Value for Latin Square Design..... 10

2.2.3 Incomplete Latin Square: The Youden Square..... 10

Factorial Design..... 12

3.1 Notation for Factorial Design..... 12

3.2 The Effect of Replication..... 13

3.3 2^f Factorial Designs..... 17

3.4 Notation for 2^f factorial Design..... 17

3.4 The YATES Scheme 20

3.5 2^3 Factorial Design 21

3.6 3^f Designs..... 23

3.6.1 YATES Scheme for 3^f Design 25

3.6.2 YATES Scheme Procedure for 3^2 Design..... 25

Interaction..... 26

Confounding..... 28

5.1 Confounding, Interaction and the ANOVA Table..... 29

5.2 Methods for Confounding..... 29

5.2.1 Table of Signs 30

5.2.2 ABD Method (Good for 2^f) 31

5.2.3 Kempthorne (Good for 2^f and 3^f Designs) 31

5.3 Confounding and the Elements of Blocks 32

5.3.1 Generating the Blocks and the elements of the Block..... 33

5.4 Confounding in 3^f 36

Fractional Factorial Design..... 37

6.1 Aliases 38

6.2 Fractional Factorial for a 2^f 39



The Design and Analysis of Engineering Experiments II
A SunCam online continuing education course

6.3	Fractional Factorial for a 3^f	41
	Random, Fixed Effect Model and Expected Mean Square.	43
7.1	Single Factor Model	43
7.2	Two-Factor Model.....	44
7.3	EMS Rules for Establishing an EMS column for ANOVA	44
7.4	Rule for Determining the Degrees of Freedom for EMS.....	46
	Nested or Hierarchical Designs	46
8.1	Testing for significance on an EMS Table.....	48
	Regression Analysis.....	48
9.1	Model Solution.....	49
9.2	Matrix Approach	50
9.3	Coefficient of Determination (R^2)	52
9.4	A Note about the Least Squares Method.....	52
	Summary	53
	REFERENCES.....	53



The Design and Analysis of Engineering Experiments II
A SunCam online continuing education course

Introduction

1.1 Role of Design of Experiments in Quality Design and Improvement

Designs of Experiments are a set of tools used to identify or screen important factors that affect a process, and to develop empirical models that characterize process behavior. It is a systematic, rigorous approach to engineering problem solving that applies principles and techniques during the data collection stage to assure the generation of valid, defensible, and supportable engineering conclusions. In addition, all of this is carried out under the constraint of minimal expenditure of engineering runs, time, and money. Also, as part of Design of Experiments, we have Response Surface Methodology, or RSM, a collection of mathematical and statistical techniques that are useful for the modeling and analysis of problems in which a response of interest is influenced by several variables and the objective is to optimize this response.

There is a need for continuous process and performance monitoring with a view towards the identification of those areas that present opportunities for product and process improvements. This makes a strong case for the need to push the quality issue farther upstream into the engineering design arena where the effects of the factors that are perceived to be important to product or process performance can be properly studied by purposefully varying or changing their levels in the experimental realm.

Specifically for process control, a crucial step is the ability to diagnose or discover the root cause, the fault that is responsible for the variation in the process/product, in order to fully understand and appreciate how best to implement process and quality improvements

Oftentimes to get to the root cause of the problem, we will need to experiment with the process, purposely changing certain factors with the hope of observing corresponding changes in the responses of the process. On the other hand, the problem could be a system problem in the sense that the process could be in control, but the variation happens to be too high, resulting in very large defect rates, and so on. This portends a fundamental problem that is not revealed easily without a comprehensive study of process performance across a range of conditions and a large number of factors. Without an organized and systematic approach to experimentation, a costly and time-consuming "random walk" approach to looking for 'root cause' or effects of change can lead to very little and perhaps nothing in terms of an enhanced knowledge of the process. The methods of design of experiments present a systematic approach that would result in an efficient and reliable procedure that would lead to better process understanding. It is important to note that the power of design of experiments can be greatly enhanced if the environment in which the experiments are conducted has been changed through variation reduction methods such as statistical process control. Statistical process control ensures a more stable process. A stable process will allow the effects of small changes in the process parameters to be more readily observed. In those cases where statistical control of a process has been established, subsequent experimentation and the associated



The Design and Analysis of Engineering Experiments II

A SunCam online continuing education course

improvement actions are more likely to result in a stable process in the future because the future is more predictable when the process is under statistical control. While statistical control of the process is not necessarily a prerequisite to drawing valid conclusions from the results of a designed experiment, it can greatly enhance the sensitivity of the experiment in the context of its ability to detect the effects of the variables.

1.2 The Role of Design of Experiments in the Design Process

A serious shortcoming of past approaches to quality has been the inability to deal rationally with the quality issue early in the product and process development life cycle. During the past several decades, it has become clear, largely through the work of Taguchi and others, that parameter selection at the early stages of product and process design can be enhanced by measuring quality by functional variation during use and by the use of experiments methods. In particular, the concept of robust design, advocated by Taguchi as part of his model for the design process, shown in Figure 2-2, has proven to be an effective tool for product and process design and improvement.

There is an important distinction to be made between testing and experimentation. While both have their rightful place, one should not serve as an alternative for the other. The Japanese have used design of experiments for parameter selection at the product and process design stage. Here, the object is to experiment with various combinations of the important design parameters to identify the particular combination(s) that optimize certain design criteria or performance measures. In the past, the West had placed a great deal of emphasis on life testing by subjecting many identical units to field conditions to determine the life expectancy of performance

Missing Values

A common problem that can ruin a good engineering design project in the process of conducting an experiment is missing values, also called missing observations. When there are missing observations, it is not possible to obtain valid measurements on some of the experimental units and so this nullifies the application of some of the techniques developed for the *ANOVA*. Missing observations occur when tools break, the machine breaks down or the operator was inattentive to collect data. In less developed countries where electricity generation and distribution are erratic, the issue of missing data may occur when the plant suddenly loses power and the backup generator malfunctions. This could lead to loss of data, the inability to continue recording/transcribing, or even the loss of already recorded data.

The otherwise straightforward analysis of randomized experiments is often complicated by the presence of missing data. Some authors suggest an approach to the problem that assumes that the data is missing at random and conditional on treatment and can be estimated by using the mean of the observations. This has been shown to be inadequate, especially when other fully observed covariates or factors are present.



The Design and Analysis of Engineering Experiments II
 A SunCam online continuing education course

2.1 Missing values for Single Factor Randomized Block Design (RBD)

An approach to solving the problem of missing values for a randomized block design is to develop an error function, which when optimized, ensures that the estimated missing value has minimum error. In other words, it should result in the optimum value of the error mean square. The optimization process requires taking the differential of the error function denoted by the quantity (Q) where Q is the modified error sum-of-squares (SS) due to the missing value and then finding the zeros of the resulting differential. It is important to note that if we had more than one missing value (say 'p' missing values), then we will take p partials and optimize with respect to the unknown parameters of interest.

To begin, let us define the model and its parameters assuming we do not have a missing value. The model is as given below. The accompanying table is a familiar data table for a treatment with several levels ('a') and blocks with several levels ('b'). In the table, the missing value is represented by the unknown variable x (not be confused with the treatment) last row, 3rd column. From there, we go on to compute the error SS.

$$y_{ijk} = \mu + \tau_i + \beta_j + \epsilon_{ij}$$

y_{ij} = observation in the i^{th} treatment and the j^{th} block

μ = overall mean

τ_i = i^{th} treatment effect, there are a treatment level, $i=1, \dots, a$,

β_j = j^{th} block effect, there are b levels of block, $j=1, \dots, b$,

ϵ_{ij} = random error

$i = 1, 2, \dots, a$ $j = 1, 2, \dots, b$
--

Table 1 layout for RBD with missing value

	Blocks						Row Totals
Treatments	y_{11}	y_{12}	y_{13}	••	••	y_{1b}	$y_{1.}$
	y_{21}	y_{22}	y_{23}	••	••	y_{2b}	••
	••	••	••	••	••	••	••
	••	••	••	••	••	••	••
	y_{a1}	y_{a2}	\underline{x}	••	••	y_{ab}	$y_{a.} + x$
Column Totals	$y_{.1}$	$y_{.2}$	$y_{.3} + x$		$y_{.b}$	$y_{..} + x$	

$$SS_{error} = \left(\sum y_{ij}^2 - CF \right) - \left(\frac{\sum y_{i.}^2}{b} - CF \right) - \left(\frac{\sum y_{.j}^2}{a} - CF \right)$$



The Design and Analysis of Engineering Experiments II
A SunCam online continuing education course

$$SS_{error} = \left(\sum y_{ij}^2 - CF \right) - \left(\frac{\sum y_{i.}^2}{b} - CF \right) - \left(\frac{\sum y_{.j}^2}{a} - CF \right)$$

but $CF = \frac{y_{..}^2}{ab}$, hence:

$$SS_{error} = \sum y_{ij}^2 - \left(\frac{\sum y_{i.}^2}{b} \right) - \left(\frac{\sum y_{.j}^2}{a} \right) + \frac{y_{..}^2}{ab}$$

For the missing value,

$$SS_{Total} = (R_0 + x^2 - CF), SS_{Treat} = \left(\frac{(y'_{i.} + x)^2}{b} + R_1 - CF \right), SS_{Block} = \left(\frac{(y'_{.j} + x)^2}{a} + R_2 - CF \right)$$

$$CF = \frac{(y'_{..} + x)^2}{ab} + R_3$$

$$SS_{Error} = (R_0 + x^2) - \left(\frac{(y'_{i.} + x)^2}{b} + R_1 \right) - \left(\frac{(y'_{.j} + x)^2}{a} + R_2 \right) + CF$$

The R's are constants, so together they form another constant, R, that is $R = R_0 + R_1 + R_2 + R_3$

$$Q = SS_{error} = x^2 - \frac{(y'_{i.} + x)^2}{b} - \frac{(y'_{.j} + x)^2}{a} + \frac{(y'_{..} + x)^2}{ab} + R$$

Differentiating with respect to x and setting equal to zero, that is: $\frac{dQ}{dx} = 0$

$$2x - \frac{2(y'_{i.} + x)}{b} - \frac{2(y'_{.j} + x)}{a} + \frac{2(y'_{..} + x)}{ab} = 0$$

Multiplying both sides by: $\frac{(ab)}{2}$, we have:

$$x(ab) - a(y'_{i.}) - ax - b(y'_{.j}) - bx + y_{..} + x = 0$$

$$x(ab - b - b + 1) = [ay'_{i.} + by'_{.j} - y_{..}]$$

$$x[b(a - 1) - (a - 1)] = x[(a - 1)(b - 1)]$$

$$x(a - 1)(b - 1) = ay'_{i.} + by'_{.j} - y_{..}$$

$$x = \frac{ay'_{i.} + by'_{.j} - y_{..}}{(a - 1)(b - 1)} = \frac{(aT - bB - G)}{(a - 1)(b - 2)}$$

<p>a=number of levels of the treatment b=number of blocks T=Total of treatment with the missing value B=Total of block with the missing value G= Grand total of all the values</p>
--

The y-primes, that is ($y'_{i.}$, $y'_{.j}$, $y'_{..}$), are the partial sums without the missing value for the affected row, column and the total and are as defined with their equivalents (T, B, G) above.

Example: An experiment to determine the amount of warping (mm) of copper plates was conducted in 4 different laboratories (Lab 1, Lab2, Lab3, Lab4) using four specimens with different percent of copper compositions (A, B, C, D)



The Design and Analysis of Engineering Experiments II
 A SunCam online continuing education course

Table 2 Layout for Randomized Block Design (RBD) with missing value

LABORATORY	SPECIMEN(Treatment)				TOTALS
	A	B	C	D	
Lab1	264	208	220	217	909
Lab2	260	231	263	226	980
Lab3	258	216	219	215	908
Lab4	241	185	225	224	875
TOTALS	1023	840	927	882	3672

Table 3 Single Factor 2-way ANOVA with One Restriction (block) on Randomization below
 ANOVA: One-Factor With Blocking

SUMMARY	Count	Sum	Average	Variance
Lab1	4	909	227.25	626.25
Lab2	4	980	245	368.6667
Lab3	4	908	227	430
Lab4	4	875	218.75	566.9167
A	4	1023	255.75	102.9167
B	4	840	210	368.6667
C	4	927	231.75	440.9167
D	4	882	220.5	28.33333

ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Rows (Labs-Block)	1468.5	3	489.5	3.253693	0.073833	3.862548
Columns (specimen)	4621.5	3	1540.5	10.23966	0.002929	3.862548
Error	1354	9	150.4444			
Total	7444	15				

Let us assume that the equipment in laboratory 4 lost power momentarily while processing so the new data configuration for specimen B, will look like that found in Tables 4 and 5.

Optimization Procedure for estimating p missing values

Develop an error function, namely Q and optimize it by taking partials (since there are more than one missing values) with respect to each missing value and set the result equal to zero. This results in p equations in p unknowns which we solve simultaneously to obtain the estimates of the missing values.

The total degree of freedom for one missing value is (n-1)-1; for two it is (n-1)-2, for three missing values the degree of freedom is (n-1)-3 and so on.



The Design and Analysis of Engineering Experiments II
 A SunCam online continuing education course

Table 4 Data table with missing value

LABORATORY	SPECIMEN(Treatment)				TOTALS
	A	B	C	D	
Lab1	264	208	220	217	909
Lab2	260	231	263	226	980
Lab3	258	216	219	215	908
Lab4	241	x	225	224	690+x
TOTALS	1023	665+x	927	882	3487+x

$$x = \frac{ay_{i\cdot} + by_{\cdot j} - y_{\cdot\cdot}}{(a-1)(b-1)} = \frac{4(690) + 4(665) - 3487}{(3)(3)} = 215 \text{ (original } y = 185)$$

Table 5 ANOVA Table for missing data

SUMMARY	Count	Sum	Average	Variance
Lab1	4	909	227.25	626.25
Lab2	4	980	245	368.6667
Lab3	4	908	227	430
Lab4	4	905	226.25	116.9167
A	4	1023	255.75	102.9167
B	4	870	217.5	93.66667
C	4	927	231.75	440.9167
D	4	882	220.5	28.33333

ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Rows	992.25	3	330.75	2.961204	0.090104	3.862548
Columns	3620.25	3	1206.75	10.80403	0.002439	3.862548
Error	1005.25	8	125.6563			
Total	5617.75	14				

2.2 A Note about the Adjusted ANOVA Table due to Missing Value.

Recall that in the analysis for the missing value we utilized 15 data points rather than 16. So, the degrees of freedom for the experiment = (N-1) = (15-1=14). After conducting ANOVA with the estimated missing value, we find in this particular case that, relative to our decision about significance, nothing has changed traumatically. The block effect was still not significant and even less so with the missing value. The treatment effect was significant in the original design and remained so with the missing value even though the F-statistic was less than before.

Question? How would the problem change if we had **two** missing values rather than one? We will proceed the same way but rather than differentiation of the function for one parameter, we will take partials with respect to two parameters, set it to zero, and solve two equations in two unknowns.



The Design and Analysis of Engineering Experiments II
A SunCam online continuing education course

2.2.2 Missing Value for Latin Square Design

$$SS_{Total} = (q^2 + x^2) - \frac{(y_{..} + x)^2}{a^2}$$

$$SS_{Row} = \frac{R^2 + (R + x)^2}{a} - \frac{(y_{..} + x)^2}{a^2} \Rightarrow (a^2 = N)$$

$$SS_{Col} = \frac{C^2 + (C + x)^2}{a} - \frac{(y_{..} + x)^2}{a^2}$$

$$SS_{Treatment} = \frac{T^2 + (T + x)^2}{a} - \frac{(y_{..} + x)^2}{a^2}$$

a=no of treatments in the experiment, same for blocks and columns
R=total of rows with missing data
C=total of Columns with missing data
T=total of treatments with missing data
y_{..}=Grand total of all observed values

$$Q = SS_{Error} = (q^2 + x^2) - \left(\frac{R^2 + (R + x)^2}{a} \right) - \left(\frac{C^2 + (C + x)^2}{a} \right) - \left(\frac{T^2 + (T + x)^2}{a} \right) + 2 \left(\frac{(y_{..} + x)^2}{a^2} \right)$$

$$\frac{dQ}{dx} = 0 \Rightarrow a^2 x - a(R + x) - a(C + x) - a(T + x) - 2(y_{..} + x)$$

$$x(a^2 - 3a + 2) = a(R + C + T) - 2y_{..}, \quad x = \frac{a(R + C + T) - 2y_{..}}{(a - 2)(a - 1)}$$

2.2.3 Incomplete Latin Square: The Youden Square

We have a Youden Square when the conditions of a Latin Square are met but only three treatments levels can be applied or are available per block because we have only 3 levels for position. Consider a machining situation where we have four machines (M=4), each with 4 positions (P=4) or heads and four possible specimens(S=4). If the machines only have **three** heads each, then the arrangement will be a Latin Square with a missing position, thus an incomplete Latin Square. Such design is called a Youden Square.

Table 6 Youden Square data

Machine(Blocks)	POSITION (heads)			T _{i..}
	1	2	3	
I	A(2)	B(1)	C(0)	3
II	D(-2)	A(2)	B(2)	2
III	B(-1)	C(-1)	D(-3)	-5
IV	C(0)	D(-4)	A(2)	-2
T _{..k}	-1	-2	1	-2
	A	B	C	D
T _{..j}	6	2	-1	-9

b= number of blocks in the experiment (b=4).

t=number of treatments in the experiment (t=4).

k= number of treatments per block (k=3).

r=number of replications of a given treatment throughout the experiment(r=3).

N= total number of observation =(b)(k) = 12.

λ= number of times each pair of treatments (say A &B) appear together throughout the experiment.



The Design and Analysis of Engineering Experiments II
A SunCam online continuing education course

$$\lambda = \frac{r(k-1)}{(t-1)} = \frac{3(2)}{3} = 2.$$

We calculate the sum-of-squares (SS):

Note: $CF = \frac{T_{..}^2}{N} = \frac{2^2}{12} = \frac{1}{3}$, $\sum \sum \sum y_{ijk}^2 = 48$

Unadjusted Block: $SS_B = \sum_{i=1}^{b=4} \frac{T_{i..}^2}{k} - CF = \frac{(3)^2 + (2)^2 + (-5)^2 + (-2)^2}{3} - \frac{1}{3} = \frac{42}{3} - \frac{1}{3} = \frac{41}{3} = 13.17$

Unadjusted Position: $SS_P = \sum_{i=1}^{b=4} \frac{T_{i..}^2}{k} - CF = \frac{(-1)^2 + (-2)^2 + (1)^2}{4} - \frac{1}{3} = \frac{6}{4} - \frac{1}{3} = 1.17$

SS Treatment (adjust for Block): $SS_{Treat} = \frac{\sum_{j=1}^{t=4} Q_j^2}{k\lambda t}$, where $Q_j = kT_{.j} - \sum_i n_{ij}T_i$.

SS Treatment (adjust for Block) $SS_{Treat} = \frac{\sum_{j=1}^{t=4} Q_j^2}{k\lambda t}$, where $Q_j = kT_{.j} - \sum_i n_{ij}T_i$.

Where $n_{ij}=1$ if treatment j appears in block i , and 0 if treatment j does not appear in block i . Note that $\sum_i n_{ij}T_i$ is merely the sum of all blocks which contain treatment j .

$Q_1=3(6)-[-2+2+3]=15$, $Q_2=3(2)-[3+2-5]=6$

$Q_3=3(-1)-[3-5-2]=-3+4=1$, $Q_4=3(-9)-[2-5-2]=-27+5=-22$

$\sum Q_j = 0 = (15 + 6 + 1 - 22)$

$$SS_{Treat} = \frac{15^2 + 6^2 + 1^2 + (-22)^2}{(3)(2)(4)} = 31.08, SS_{Error} = 47.67 - 13.67 - 31.08 - 1.17 = 1.75$$

Table 7 ANOVA Table for the Youden Square data

Source	SS	df	MS	F
Mach(unadjust.)	13.67	3	-	
Treat (Adjust)	31.08	3	10.4	17.86
Position	1.17	2	0.58	1
Error(Residual)	1.75	3	0.58	
Total	47.67	11		

Based on the ANOVA, the treatment effect was significant ($F_{0.05,3,3} = 9.2766$). The position effect was not significant. We did not test for the block effect because it was not adjusted. We only test for effects that have been adjusted or those that do not need adjustment. However for symmetrical designs where $t=b$ such as we have here, we can also adjust the SS for the block as we did for treatment but **both cannot** be tested at the same time. There was no need to adjust for position because every position was present in every block and every treatment.



The Design and Analysis of Engineering Experiments II

A SunCam online continuing education course

Factorial Design

In factorial experiments, the factors (treatments) are typically quantitative (note that we said ‘typically’ because there are factors that are qualitative, such as depth, when denominated as low, medium or high) and quantitative when the actual numerical measurement of depth is used. For a quantitative factor, the difference between any two levels is an interval or ratio and such interval is not only measurable and quantifiable but also has meaning. For example, the difference between the two temperature levels of 50°C and 40°C are quantifiable. However, in the case of position, the difference between position 1 and position 2 may be understood but cannot be quantified. This becomes especially important when it is desired to establish a functional relationship or predictability function between the factor and the response. Restrictions (blocks) are not factors. They are considered nuisances that cloud the response.

In the cases we have encountered so far, we have been concerned only with one factor and the effect of such a factor on a measured response variable. Every other thing we have looked at was with respect to restrictions and randomization. The idea was to identify and remove the effect of the restriction (block or nuisance) so as not to cloud or confound the response and the error mean square. Several different designs of this type were examined. However, each type only focused on restrictions on randomization, but still the focus was on a single factor.

By contrast, a factorial experiment is one that involves more than one factor (treatment). In most cases, these factors or treatments may have some functional relationship that defines their behavior relative to the response variable. Also, in a factorial **experiment, all levels of a given factor are combined with all levels of every other factor** in the experiment. Factorial designs are most efficient in those situations where there are two or more factors, thus reducing cost. By factorial design, we mean that for each complete trial or realization of the experiment all possible levels of the factors are run and data obtained. For example, if factor A has ‘a’ levels and factor B has ‘b’ levels; we will have a total of ab ($a \times b$) treatment combinations. Thus, one of the major benefits of a factorial design is that it allows the effect of several factors and in some cases, the interactions among those factors to be determined with the same number of experimental trials needed for the one-factor at a time design, thus reducing the cost of experimentation. With two or more factors, the traditional approach has been to hold one or more factors constant, while the other factor is varied through its different levels. After taking the readings, another factor is chosen and the process repeated until all the factors have been exhausted. This requires a lot of experimental trials. However, with factorial experiments, all possible combinations are run at the same time, thus reducing the cost of the experiment.

3.1 Notation for Factorial Design

If we have A1, A2, A3 as the levels of factor A; B1, B2, B3, B4 as levels of factor B; and C1, C2, C3 as the levels of factor C, then we have a $3 \times 4 \times 3$ factorial design with 36 data points.



The Design and Analysis of Engineering Experiments II
 A SunCam online continuing education course

The special case is where we have 2^f (f factors at 2 levels) or 3^f (f factors at 3 levels). These are the most common type of factorial designs. Higher levels or higher order designs are more complicated and are usually avoided. In these designs the factors are completely randomized and the levels are considered fixed. We will examine cases much later where the levels are random.

3.2 The Effect of Replication

A typical functional relationship in a $2 \times 2 = 4$ factorial design features two factors at two levels. The model representing this design looks as follows:

$$y_{ij} = \mu + A_i + B_j + (AB)_{ij} + \varepsilon_{ij} \text{ (Two-Way without replication)}$$

The above model shows that for a 2-factor design without replication, the effect due to interaction is indistinguishable from the effect due to error. This presents a problem in the analysis, especially when it is desired to estimate the interaction effect. A way to solve this problem is to carry-out replication for each data point. That way it would be possible to independently estimate the error mean square and the interaction effect. The model with replication will look like:

$$y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + \varepsilon_{k(ij)}$$

Example: (Two-way with replication). The following data for two factors, namely temperature (T-3 levels) and accelerator --% of calcium chloride (A-3 levels) with 2 replications and their effect on the cure time of concrete. The cure time is the measured output.

The model is as follows: $y_{ijk} = \mu + C_i + A_j + CA_{ij} + \varepsilon_{k(ij)}$,

Where, $C_i=0$ for all i, $A_j=0$ for all j, $CA_{ij} =$ for all (ij) combinations

Table 8 -Data Table for Factorial Design with Replication

Temp	Accelerators(% Cacl2)			T _{i..}
	40%	50%	60%	
40	-1	3	-3	-7
	(-5)	(0)	(-1)	
50	-4	3	1	-4
	1	2	-1	
60	(-4)	(1)	(-1)	-9
	-5	-1	0	
T _{.j.}	0	3	-4	-20
	(-2)	(-1)	(-6)	
	-2	-4	-2	
	-11	0	-9	



The Design and Analysis of Engineering Experiments II
 A SunCam online continuing education course

Preliminary Calculations

$$\sum \sum \sum y_{ijk}^2 = (-1)^2 + (-4)^2 + (3)^2 + (3)^2 + (-3)^2 + (1)^2 + (1)^2 + \dots + (-2)^2 = 126$$

$$CF = (-20)^2 / 18 = 22.22,$$

$$Cell : \frac{\sum \sum T_{ij\cdot}^2}{2} = \frac{(-5)^2 + (0)^2 + (-1)^2 + (-4)^2 + (-1)^2 + (-1)^2 + \dots + (-6)^2}{2} = 44$$

$$Temp : \frac{\sum T_{i\cdot\cdot}^2}{6} = \frac{(-7)^2 + (-4)^2 + (-9)^2}{6} = 24.33, \quad Accel : \frac{\sum T_{\cdot j\cdot}^2}{6} = \frac{(-11)^2 + (0)^2 + (-9)^2}{6} = 33.67$$

$$SSTotal = 126 - 22.22 = 103.8, \quad SS_{cell} = 44 - 22.22 = 21.8,$$

$$SSTemp = 24.33 - 22.22 = 2.1, \quad SS_{Accel} = 33.67 - 22.22 = 11.45$$

$$SS(interaction) = SS_{cell} - SSTemp - SS_{Accel} = 8.2$$

$$SS Error = SSTotal - SS(interaction) = 103.8 - 8.2 = 95.6$$

Table 9 -Data Table for Factorial Design with Replication

Source	SS	df	MS	F
Temp(C)	2.1	2	1.05	< 1
Accel (A)	11.45	2	5.75	< 1
CA _{ij}	8.2	4 (2x2)	2.05	< 1
ε _{k(ij)}	82	9	9.11	
Total	47.67	17		

Example of a 3-way Factorial Design with Replication.

The factors are Concentration (40%, 50%, 60%), Temperature (50°C, 60°C), and Catalyst (A, B)

$$y_{ijk} = \mu + C_i + M_j + CM_{ij} + T_k + CT_{ik} + MT_{jk} + CMT_{ijk} + \varepsilon_{ijk}$$

As can be seen from the model equation, CMT_{ijk} is confounded with ε_{ijk}.

We can now compute the sum-of-squares of the different components of the ANOVA table. Before we do that, it is important that we discuss the method of computing the sum-of-squares especially the interaction when we have three or more factors. We develop two-way tables for pairs of factors to facilitate the computation of the first order interaction,. Such tables make it easy to compute the sum-of-squares of the individual factors and the sum-of-squares for their interaction, as we will demonstrate shortly.



The Design and Analysis of Engineering Experiments II
 A SunCam online continuing education course

Table 10 Data table for 3-way (3 factors) with no Replication

CATA(M)	TEMP(T)	CONC(C)		
		40%	50%	60%
A	50°C	1	7	-1
	60°C	-2	8	-3
B	50°C	-120	7	88
	60°C	-124	5	92

Table 11 Two-way tables for Computing main effects and interaction

		CATA (M)			
		A	B		
T•jk	TEMP	50 °C	7	-25	(M x T)
	(I)	60 °C	3	-27	
	T•j•		10	-52	-42
		CONC (C)			
		40%	50%	60%	T••k
Ti•k	TEMP	50	-119	14	87
	(I)	60	-126	13	89
	(CxT)		-245	27	176
					-42
		CONC(C)			
		40%	50%	60%	
Tij•	CATA	A	-1	15	-4
	(M)	B	-244	12	180
	Ti••		-245	27	176
					-42



The Design and Analysis of Engineering Experiments II
A SunCam online continuing education course

$$CF = (-42)^2 / 12 = 147$$

$$SS(Total) = \sum \sum \sum y_{ijk}^2 = 1^2 + 7^2 + (-1)^2 + (-2)^2 + 8^2 + (-3)^2 + \dots + 92^2 - CF = 46186 - 147 = 46039$$

$$SS(M) = \sum T_{.j\bullet}^2 = \frac{10^2 + (-52)^2}{6} - CF = 467.333 - 147 = 320.333 \text{ (from the } T_{.jk} \text{ table)}$$

$$SS(T) = \sum T_{i\bullet k}^2 = \frac{(-18)^2 + (-24)^2}{6} - CF = 150 - 147 = 3 \text{ (from the } T_{i\bullet k} \text{ table)}$$

$$SS(C) = \sum T_{i\bullet\bullet}^2 = \frac{(-245)^2 + (27)^2 + (176)^2}{4} - CF = 22932.5 - 147 = 22785.5 \text{ (from the } T_{ij\bullet} \text{ table)}$$

We now compute the interactions based on the two-way tables

$$SS(M \times T) = \sum \sum T_{.jk}^2 = \frac{(7)^2 + (-25)^2 + (3)^2 + (-27)^2}{3} - CF - SS(M) - SS(T)$$

$$= 470.67 - 147 - 320.33 - 3.0 = 0.34$$

$$SS(C \times T) = \sum \sum T_{i\bullet k}^2 = \frac{(-119)^2 + (14)^2 + (87)^2 + (-126)^2 + (13)^2 + (89)^2}{2} - CF - SS(C) - SS(T)$$

$$= 22946 - 147 - 22785.5 - 3 = 10.5$$

$$SS(C \times M) = \sum \sum T_{ij\bullet}^2 = \frac{(-1)^2 + 14^2 + (-4)^2 + (-244)^2 + 12^2 + 176^2}{2} - CF - SS(C) - SS(M)$$

$$= 46161 - 147 - 22785.5 - 320.33 = 22908.17$$

With these sum-of-squares, we can now establish the ANOVA table and test for significance. Please note that because there is no replication, the highest order interaction (C x M x T) is confounded with the error sum-of-squares and so we are unable to tell them apart. This is not a problem because in most experiments (especially higher order experiments), the physical significance of the highest order interaction is difficult to explain so it is usually confounded or lumped with the error term to enhance or improve the test. It is important to note that when there is a need to add the two sums of squares, the degrees of freedom are also added to the degrees of freedom of the error term to give both a new SS and df for the error. If we had replication of $r = 2$, then,

df total = $(24-1)=23$, df CMT = $(2)(1)(1) = 2$, df CM=2, df CT=2, df MT=1, then, df error=(df total)-(df C)-(df M)-(df T)-(df CM)-(df CT)-(df MT)-(df CMT)= $23-11=12$

Table 12 ANOVA Table for the 3-way (3 Factor) Design with Replication

Source	SS	df	MS	F	Sig	F(Table)
C _i	22785.5	2	11393	2042	***	F(2,2,0.01)=99.0
M _j	320.33	1	320.33	57.41	ns	F(1,2,0.01)=4999.5
T _k	3.00	1	3	0.538	ns	F(1,2,0.01)=4999.5
CM _{ij}	22908.17	2 (2x1)	11454	2053	***	F(2,2,0.01)=99.0
CT _{ik}	10.5	2(2x1)	5.25	0.941	ns	F(2,2,0.01)=99.0
MT _{jk}	0.34	1(1x1)	0.34	0.061	ns	F(1,2,0.01)=4999.5
ε _{κ(ij)}	11.16	2	5.58			
Total	46039	11				



The Design and Analysis of Engineering Experiments II
A SunCam online continuing education course

3.3 2^f Factorial Designs

The 2^f are designs with n factors where each factor has just two levels. The two levels are usually at the extremes and so this type of design is used to screen or to carry out investigation about the factors and their levels. By screening, we are referring to the process of exploring several factors and the levels of such factors with the hope of selecting those important factors and their levels for the response under consideration.

We will start with the basic case of this type of design, namely 2² or the 2 x 2 factorial with complete randomization and no replication. Since we do not have replication, this means that the interaction effect is confounded or confused with the error. For notation purposes for 2^f, we will be using a mnemonic device that was developed by YATESs and others to make the notation, ordering, representation, and analysis of the experiment easy to implement. As we will see later, for a factor with two levels, the most we will be able to estimate in terms of the functional relationship is the linear component. As we may recall, with two points, we can locate a line. In the same vein, with three points we potentially have a curvature.

3.4 Notation for 2^f factorial Design

The notation which has been widely used is as follows. Consider, as in our case, two factors A and B. Let 0 and 1 denote the two levels of a 2^f design. So A₀B₀ is used to denote that both factors are at their low levels. A₁B₀ denotes that A is at high level and B is at low level, A₀B₁ denotes A low and B high, and A₁B₁ denotes both at the high levels. Given this notation, we then come up with another response notation for the factors at these levels. The response notation is quite intuitive, as we will see momentarily.

- When both are at their low levels (A₀B₀), we represent the response as: (1),
- For data point A₁B₀, we represent it as the response as ‘a’.
- Similarly for A₀B₁:b,
- and for A₁B₁:ab.

This is done in order, starting with all factors at their low levels, then the first factor, then the second factor. If a third factor is introduced, we first exhaust the sequence for the first two factors before we introduce the third and so on in that order.

2 Factors at 2 levels	
00	(1)
10	a
01	b
11	ab

3-Factors at 2 levels	
000	(1)
100	a
010	b
110	ab
001	c
101	ac
011	bc
111	abc



The Design and Analysis of Engineering Experiments II
 A SunCam online continuing education course

Effects of a factor.

To determine the effects of a factor, say A, we examine what happens to A when B is low and also when B is high. The average of these two will give us the average effect of A, similarly for B.

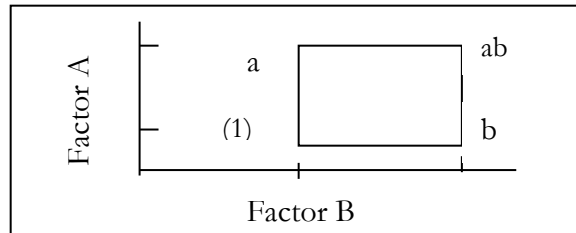


Figure 1 Two Factor (A, B) Factorial Design at 2 levels

From figure 1, the effect of A At the low B is [(a-(1))]. The effect at high B, is [ab-b]. Therefore, the average effect of A is:

$$= \frac{1}{2} [(a - (1)) + (ab - b)] = \frac{1}{2} [-(1) + a - b + ab]$$

An intuitive way to look at this is the following. Where A exists, you have a positive sign (+), where it does not we have a negative sign (-). Using this rule of thumb, we can develop the effect of B as follows:

$$B = \frac{1}{2} [-(1) - a + b + ab]. \text{ Similarly, for AB: } \frac{1}{2} [+ (1) - a - b + ab]$$

Another thing to notice at this stage is that the signs are equal. To summarize:

$$A = \frac{1}{2} [-(1) + a - b + ab] \Rightarrow 2A = [-(1) + a - b + ab]$$

$$B = \frac{1}{2} [-(1) - a + b + ab] \Rightarrow 2B = [-(1) - a + b + ab]$$

$$AB = \frac{1}{2} [+ (1) - a - b + ab] \Rightarrow 2AB = [+ (1) - a - b + ab]$$

In addition to the effect, we also have contrasts as follows:

$$2A = [-(1) + a - b + ab] \Rightarrow -1 + 1 - 1 + 1 = 0$$

$$2B = [-(1) - a + b + ab] \Rightarrow -1 - 1 + 1 + 1 = 0$$

$$2AB = [+ (1) - a - b + ab] \Rightarrow +1 - 1 - 1 + 1 = 0$$

Each of 2A, 2B, 2AB is a contrast and together they form a set of orthogonal contrasts. To demonstrate that for two of the contrasts A, AB; [-1-1+1+1] = 0. Additionally, because there are two levels for a 2f design, we can also represent the levels with +1 and -1, depending on whether the effect is present or not present. The low level always assumes the negative sign for each effect. Once this is done, we can also determine the coefficient of the interaction (in this case AB) by multiplying the proper corresponding signs of the effects. We can also use the table of signs (table 13) to determine the Main effect and the interaction effects 2³. We will show proof of concept by



The Design and Analysis of Engineering Experiments II
A SunCam online continuing education course

looking at the table of signs for 2^3 (8 total points). For each interaction, we get the sign of the effect by multiplying by the signs of the factors that make up the interaction.

Table 13 Table of Signs for 2^f Factorial Design

Treatment Combinations	Effects							
	Response	A	B	AB	C	AC	BC	ABC
(1)	+	-	-	+	-	+	+	-
a	+	+	-	-	-	-	+	+
b	+	-	+	-	-	+	-	+
ab	+	+	+	+	-	-	-	-
c	+	-	-	+	+	-	-	+
ac	+	+	-	-	+	+	-	-
bc	+	-	+	-	+	-	+	-
abc	+	+	+	+	+	+	+	+

Consider the following 2x2 design:

Table 14 New Notation and data Representation for 2^2 Design

	Factor A		
Factor B	A1	A2	TOTAL
B1	30 (1)	40 a	70
B2	0 b	20 ab	20
TOTAL	30	40	

$$A_0B_0=30=(1), A_1B_0=a=40, A_0B_1=b=0, A_1B_1=ab=20$$

From the normal ANOVA, approach, with replication $r=1$ in this example.

$$CF = (90)^2 / 4 = 2025, \sum \sum y_{ij}^2 = 30^2 + 40^2 + 0 + 20^2 = 2900, SS_{Total} = 2900 - CF = 875$$

$$SS_B = \frac{70^2 + 20^2}{2} - CF = 625, SS_A = \frac{30^2 + 40^2}{2} - CF = 225, SS_{AB} = 875 - 225 - 625 = 25$$

Using the new Approach, the effects are: Define the following:

$$Contrast = r \cdot 2^{f-1} (Effect)$$

$$Effect = \frac{1}{r(2^{f-1})} (Contrast), SS_{Effect} = \frac{(Contrast)^2}{r \cdot 2^f}$$

$$A = \frac{1}{2} [-30 + 40 + 20] = 15, B = \frac{1}{2} [-30 - 40 + 0 + 20] = -25, AB = \frac{1}{2} [30 - 40 - 0 + 20] = 5$$

The Contrasts are: $2A=30, 2B=-50, 2AB=10$

$$SS_A = \frac{(30)^2}{1 \times 2^2} = \frac{900}{4} = 225, SS_B = \frac{(-50)^2}{1 \times 2^2} = 625, SS_{AB} = \frac{(10)^2}{1 \times 2^2} = 25$$



The Design and Analysis of Engineering Experiments II
A SunCam online continuing education course

3.4 The YATES Scheme

Using the table of signs, we will introduce yet another way of computing the sum-of-squares and that is the YATES scheme or YATES method. In the Yates methods, the treatments are added in pairs all the way through. That forms half of the elements in a given iteration. Then the same set of numbers is again manipulated by subtracting the first elements from the second and continuing to till the list is exhausted. This second operation then fills the remainder of the elements of the iteration. This continues for the next iteration and so on until the number of iterations equals the number of factors. In our example, there are two factors, so the number of iterations is two or two columns. At the end, the third last column is used to indicate the sum-of-squares. The sum-of-squares is obtained by dividing the results of the f^{th} column (or second column in this example) by the divisor defined as: $r \times 2^f = N$ which essentially is the number of data points in the experiment.

Table 15 YATES Scheme Solution for 2^2 Design

Treat Com.	Resp.	I	II	SS= (Col II) ² /($r \times 2^f$)
(1)	30	70	90	2025=CF
a	40	20	30	225 =SS _A
b	0	10	-50	625=SS _B
ab	20	20	10	25=SS _{AB}

The following will help explain how the **YATES Scheme** works. Starting with the response column:

1. First Iteration for Column I

Starting with the Response Column

- Add 30 to 40 =70
- Add 0 to 20 =20

Done with the addition.

Next is subtraction

- Subtract 30 from 40=10
- Subtract 0 from 20=20

2. Second Iteration for Column II

Starting with Column I

- Add 70 to 20=90
- Add 10 to 20 =30

Done with addition

- Subtract 70 from 20=-50
- Subtract 10 from 20=10

3. End after f iterations. In our case $f=2$ the number of factors

4. Now divide elements in the last column by the divisor ($r \times 2^f$) to give the sum-of-squares



The Design and Analysis of Engineering Experiments II
 A SunCam online continuing education course

3.5 2³ Factorial Design

Consider a chemical process where the yield is of concern. Three factors, Temperature (A), Humidity (B), and % Catalyst (C) are of concern. The experiment was run with 3 replications (r = 3). The coded data is as shown. The model is:

$$y_{ijk} = \mu + A_i + B_j + AB_{ij} + C_k + AC_{ik} + BC_{jk} + ABC_{ijk} + \varepsilon_{(ijk)l}$$

Table 16 Data Table for 2³ factorial Design with replication

	A1				A2				Ti•
	B1		B2		B1		B2		
C1	-2		3		-3		-1		5
	-3	-1	3	5	-1	-3	2	4	
	4	(1)	-1	b	1	a	3	ab	
C2	-1		2		2		0		13
	0	6	-2	1	-3	3	2	3	
	7	c	1	bc	4	ac	1	abc	
T•j	5		6		0		7		T••=18
	B1 B2		A1 A2		C2 C2				
	5 13		11 7		5 13				

Table 17 The three 2-way tables for the 2³ Design

		B1	B2	
T•jk	c1	-4	9	5
	c2	9	4	13
	T•j•	5	13	18
		A1	A2	T••k
Ti•k	C1	4	1	5
	C2	7	6	13
		11	7	18
		A1	A2	
Tij•	B1	5	0	5
	B2	6	7	13
	Ti••	11	7	18



The Design and Analysis of Engineering Experiments II
A SunCam online continuing education course

Treat	Resp	I	II	III	SS	
(1)	-1	-4	5	18	13.50	CF
a	-3	9	13	-4	0.67	SS(A)
B	5	9	-3	8	2.67	SS(B)
ab	4	4	-1	6	1.50	SS(AB)
C	6	-2	13	8	2.67	SS(C)
ac	3	-1	-5	2	0.17	SS(AC)
bc	1	-3	1	-18	13.50	SS(BC)
abc	3	2	5	4	0.67	SS(ABC)

As a check, we will use the traditional method to compute the SS for the main effects and the interactions to confirm that, indeed, the YATES Scheme in this case of 2^3 indeed works.

$$CF = 18^2 / 24 = 13.50$$

$$SS_A = \frac{11^2 + 7^2}{12} - CF = 14.167 - 13.5 = 0.67$$

$$SS_B = \frac{5^2 + 13^2}{12} - CF = 16.167 - 13.5 = 2.67$$

$$SS_C = \frac{5^2 + 13^2}{12} - CF = 16.167 - 13.5 = 2.67$$

We can also set up the 2-way tables (3 in all) to see if we can compute the interactions the usual way and compare with the YATES Scheme

$$SS(B \times C) = \sum \sum T_{jk}^2 = \frac{(-4)^2 + (9)^2 + (9)^2 + (4)^2}{6} - CF - SS(B) - SS(C)$$

$$= 32.33 - 13.5 - 2.67 - 2.67 = 13.50$$

$$SS(A \times C) = \sum \sum T_{ik}^2 = \frac{(4)^2 + (1)^2 + (7)^2 + (6)^2}{6} - CF - SS(A) - SS(C)$$

$$= 17.00 - 13.5 - 0.67 - 2.67 = 0.17$$

$$SS(A \times B) = \sum \sum T_{ij}^2 = \frac{(5)^2 + (0)^2 + (6)^2 + (7)^2}{6} - CF - SS(A) - SS(B)$$

$$= 18.34 - 13.5 - 0.66 - 2.67 = 1.50$$

$$SS(Total) = (-2)^2 + (-3)^2 + (4)^2 + (3)^3 + (3)^2 + (-1)^2 + \dots + (0)^2 + (1)^2 + (2)^2 - CF$$

$$166 - 13.5 = 152.5$$

$$SS(Cell) = \frac{(-1)^2 + (5)^2 + (-3)^2 + (4)^2 + (6)^2 + \dots + (3)^2}{3} - CF = 31.33 - 13.50 = 17.83$$

$$SS(A \times B \times C) = SS(Cell) - SS(AB) - SS(AC) - SS(BC) - SS(A) - SS(B) - SS(C)$$

$$21.84 - 1.5 - 0.17 - 13.5 - 0.67 - 2.67 - 2.67 = 0.66$$

It is clear from the computations that the conventional method agrees with the YATES scheme.



The Design and Analysis of Engineering Experiments II

A SunCam online continuing education course

3.6 3^f Designs

3^f designs are very powerful and yet complicated designs. The motivation for a 3² design is the need to dig deeper to understand the functional relationships between the factors and the response. However, we will limit our work to the 3² design, namely 2 factors at 3 levels. From basic engineering principles, it is known that if you desire a linear function or linear relationship, you will need at least two data points. Similarly, if there is a need for a quadratic or curvilinear relationship one would need at least three data points. However, for most basic engineering designs, the quadratic and the cubic are mostly used because they are more easily explained. This does not obviate the fact that when there is need for higher order representation based on theoretical considerations (such as in space systems) that a cubic and higher order design may not be used. Most engineers will confess that when a relationship is beyond the cubic, quartic or quintic, it is not only very difficult to visualize but also to explain to clients and colleagues what such relationship looks like in practical terms. Again, it is much easier to explain to a client that there exists a quadratic relationship or even a cubic than it is to explain a quartic or quintic representation or relationship.

As we indicated earlier, 2^f designs are called screening designs because we employ them to understand the limits of the response. Typically, the two levels for a 2^f factorial represent the lower and upper limits of the response. Once we have that information, we have a good idea about the behavior of the process. In 3^f design, we add another level to have a more comprehensive idea of the relationship. One of the benefits of a 3^f design is that even if the linear relationship is what is of interest, we know that from basic design principles, a quadratic function also includes a linear function so going up one more level would enable us understand the relationship surface more comprehensively and still cater to the need of what is really of interest. The only drawback to this approach is of course the problem of over fitting.

The over fitting issue comes about because each parameter (or term) that is introduced into the model uses up one degree of freedom. As an example, suppose we have 4 data points and we introduce 3 parameters. Because of the constant term in the model, there will only be three degrees of freedom left to be shared among the parameters including the error term. In this scenario, there will be no degree of freedom left to use for the error term. A good way to make sure that we avoid this problem is to use replication. With replication there is always going to be enough degrees of freedom to estimate of pure error.

Going back to the 3^f design, the 3 levels are represented by 0, 1 and 2. To apply the method of the YATES Scheme, we will use the third order polynomial ($k = 3$) to compute the column values on the YATES table and then the sum-of-squares. Before we do that let's look at a conventional way to compute the sum-of-squares (SS) and the *ANOVA* table for 3^f design.



The Design and Analysis of Engineering Experiments II
A SunCam online continuing education course

Example: We are to consider the effect of two factors (Temperature and Humidity) on the yield of a process. The data is as follows:

Table 19 Data Table for 3² Factorial Design with replication

		TEMPERATURE							
		COLD		AMBIENT		HOT		T _{i..}	
HUMIDITY		0.8		1.5		2.5			
	50%		3.6		4.7		6.7	15.0	
		2.8		3.2		4.2			
		1.0		1.6		1.8			
	70%		2.6		3.4		2.8	8.8	
		1.6		1.8		1.0			
		2.0		1.5		2.5			
	90%		4.2		2.3		6.5	13	
		2.2		0.8		4.0			
	T _{.j.}	10.4		10.4		16		36.8	T _{..}

$$\text{Model: } y_{ijk} = \mu + H_i + T_j + HT_{ij} + \varepsilon_k(ij)$$

$$\sum \sum \sum y_{ijk}^2 = 92.44, \quad CF = \frac{(36.8)^2}{18} = 75.24,$$

$$SS(\text{Total}) = 92.44 - 75.24 = 17.20$$

$$SS(\text{Cell}) = \frac{(3.6)^2 + (4.7)^2 + \dots + (6.5)^2}{2} - CF = 85.64 - 75.24 = 10.40$$

$$SS(H) = \frac{\sum T_{i..}}{6} - CF = \frac{(15)^2 + (8.8)^2 + (13)^2}{6} - 75.24 = 78.57 - 75.24 = 3.33$$

$$SS(T) = \frac{\sum T_{.j.}}{6} - CF = \frac{(10.4)^2 + (10.4)^2 + (16.0)^2}{6} - 75.24 = 78.72 - 75.24 = 3.48$$

$$SS(HXT) = SS(\text{Cell}) - SS(H) - SS(T) = 10.40 - 3.33 - 3.48 = 3.59$$

$$SS(\text{Error}) = SS(\text{Total}) - SS(\text{Cell}) = 17.20 - 10.40 = 6.8$$

Table 20 ANOVA for the 3 2 Design

Source	SS	Df	MS	F(computed)	F (table)	Sig
H _i	3.33	2	1.67	2.20	F(0.01, 2,9)=3.01	n.s.
T _j	3.48	2	1.79	2.29	F(0.01, 2,9)=3.01	n.s.
HT _{ij}	3.59	4	0.90	1.19	F(0.01, 4,9)=2.69	n.s.
ε _{k(ij)}	6.80	9	0.76			
Total	17.20	17				

Based on the results from the ANOVA table, none of the factors appear statistically significant at α equals at 1% level



The Design and Analysis of Engineering Experiments II
A SunCam online continuing education course

3.6.1 YATES Scheme for 3^f Design

For a 3^f design, the minimum number of data points is 9, which represents a 3² design. If you recall in the case of the 2^f design, the operations performed to get the sum-of-squares (SS) on the YATES table was carried out in pairs. So, in the case of the 3^f, the operations will be carried out in triples or in threes. The following procedure will be used to obtain the SS on the YATES for 3^f.

Before we go further, we want to re-write the model to reflect the fact that in a 3^f design we have constant, linear and quadratic components. Note that, in general, we can do this for any factor. However, in practical terms, it makes sense only for quantitative factors since for qualitative factors, the differences in the levels are not ratios or intervals.

$$y_{ijk} = \mu + T_1 + T_2 + H_1 + T_1H_1 + T_2H_1 + H_2 + T_1H_2 + T_2H_2 + \varepsilon_{k(ij)}$$

3.6.2 YATES Scheme Procedure for 3² Design

1. Obtain the orthogonal polynomial for k=3 (that is, the number of levels for a 3^f design). As we noted previously, the maximum number of orthogonal polynomial for k levels is k-1, so in the case of 3 levels, we have 2 orthogonal polynomials as follows:

$$\begin{bmatrix} -1 & 0 & 1 \\ 1 & -2 & 1 \end{bmatrix}$$

These represent the linear and quadratic components, so we add the constant component to give us the complete polynomial structure. The result will look as follows:

$$\begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 1 & -2 & 1 \end{bmatrix}$$

With this we can then compute the YATES scheme table entries to yield the SS. As we indicated, this will be accomplished in triplets following the 3 x 3 orthogonal matrix.

2. Compute the table elements as follows

a). First Iteration:

Add the values as triplets, namely, first three values, the second three, and then the last three in the case of 3^f. This is represented by (1 1 1). In the case of 3², there will be only three sets of this type of addition because there are only 9 values

b). Second Iteration:

Subtract the 1st value from the 3rd for the first three values, the same for the second set of three values and same for the third set of 3 values all the way through. This is represented by (-1 0 1).

c). Third iteration:

Add the 1st and 3rd values and subtract from the result 2x the middle value for the first set of three values, same for the second, and same for the third set of three values. This represented by (1 -2 1).



The Design and Analysis of Engineering Experiments II
 A SunCam online continuing education course

Similar to 2^f design, repeat the procedure f-times (2 times in this case), where f represents the number of factors. **Note** that we performed each operation 3 times because we have a 3^2 design or nine data points. If we had a 3^3 design, then we will have 27 data points so we have to perform each set of operations nine times.

3. Compute the Divisor: The divisor is given by: $Divisor = 2^m 3^{f-p} r$, Where,
 M = number of letter involved in the interaction (m=0, 1 or 2), for T_1H_1 , m=2 (due to T, H)
 P = number of linear terms in the model (for T_1H_1 , p=1)
 F = number of factors (for this example, f=2)
 R = number of replications (for this example r=2)

For example based on this 3^2 design

$$T_2H_1; p=1, m=2, 2^m 3^{f-p} r = 2^2 3^1 2 = 24, T_2; p=0, m=1, 2^m 3^{f-p} r = 2^1 3^2 2 = 36$$

$$T_2H_2; p=0, m=2, 2^m 3^{f-p} r = 2^2 3^2 2 = 72$$

Table 21 YATES Scheme for 3^2 Design Using data from Previous Example

Treatment Comb	Resp.	I	II	Effect	Divisor = $2^m 3^{f-p} r$	SS=Col (II) ² /Divisor
(1) 00	3.6	15.0	36.6	CF	$2^0 3^2 2 = 18$	75.24
t ₁ 10	4.7	8.8	5.6	T ₁ (linear)	$2^1 3^1 2 = 12$	2.61
t ₂ 20	6.7	13.0	5.6	T ₂ (quadratic)	$2^1 3^2 2 = 36$	0.87
h ₁ 01	2.6	3.1	-2.0	H ₁	$2^1 3^1 2 = 12$	0.33
t ₁ h ₁ 11	3.4	0.2	-0.8	T ₁ H ₁	$2^2 3^0 2 = 8$	0.08
t ₂ h ₁ 21	2.8	2.3	5.2	T ₂ H ₁	$2^2 3^1 2 = 24$	1.12
h ₂ 02	4.2	0.9	10.4	H ₂	$2^1 3^2 2 = 36$	3.00
t ₁ h ₂ 12	2.3	-1.4	5.0	T ₁ H ₂	$2^2 3^1 2 = 24$	1.04
t ₂ h ₂ 22	6.5	6.1	9.8	T ₂ H ₂	$2^2 3^2 2 = 72$	1.33

We can now re-do the ANOVA table with the linear and quadratic components and show their significance. Please note: $SS(T) = SS(T_1) + SS(T_2) = 2.61 + 0.87 = 3.48$

$$SS(T \times H) = SS(T_1H_1) + SS(T_2H_1) + SS(T_1H_2) + SS(T_2H_2) = 0.08 + 1.12 + 1.04 + 1.33 = 3.57$$

$$SS(H) = SS(H_1) + SS(H_2) = 0.33 + 3.00 = 3.33$$

Interaction

When several factors are involved in an experiment, the effect of one of the independent variables on the response may depend on the level of the other factors in the experiment. When the difference in the response between the levels of one factor is not the same at all levels of the other factors, we have interaction. Thus, if the effect of a given factor on the response variable is



The Design and Analysis of Engineering Experiments II
 A SunCam online continuing education course

dependent at the level of the other factors then we have interaction. If a change in a factor (A) produces a different change in the response variable (Y) at one level of another factor (B) than at other levels of factor (B), we have interaction.

Example: Consider three factors Concentration (C: 50%, 60%), Temperature (T: 60C, 50C), and Catalyst (M: 1, 2).

Figure 2: At Low Temp (T); $C=87-14=73$, At High Temp (T); $C= 120-50=70$.

Interaction CT is the average difference of change in effect of C= $(70-73)/2=-1.5$ (quite small)

Figure 3: At Low Catalyst (M); $C=-4-55=-59$ At High Catalyst (M); $C= 180-12= 168$.

Interaction CM is the average difference of change in effect of $C=(168-(-59))/2=163.55$ (This is quite large)

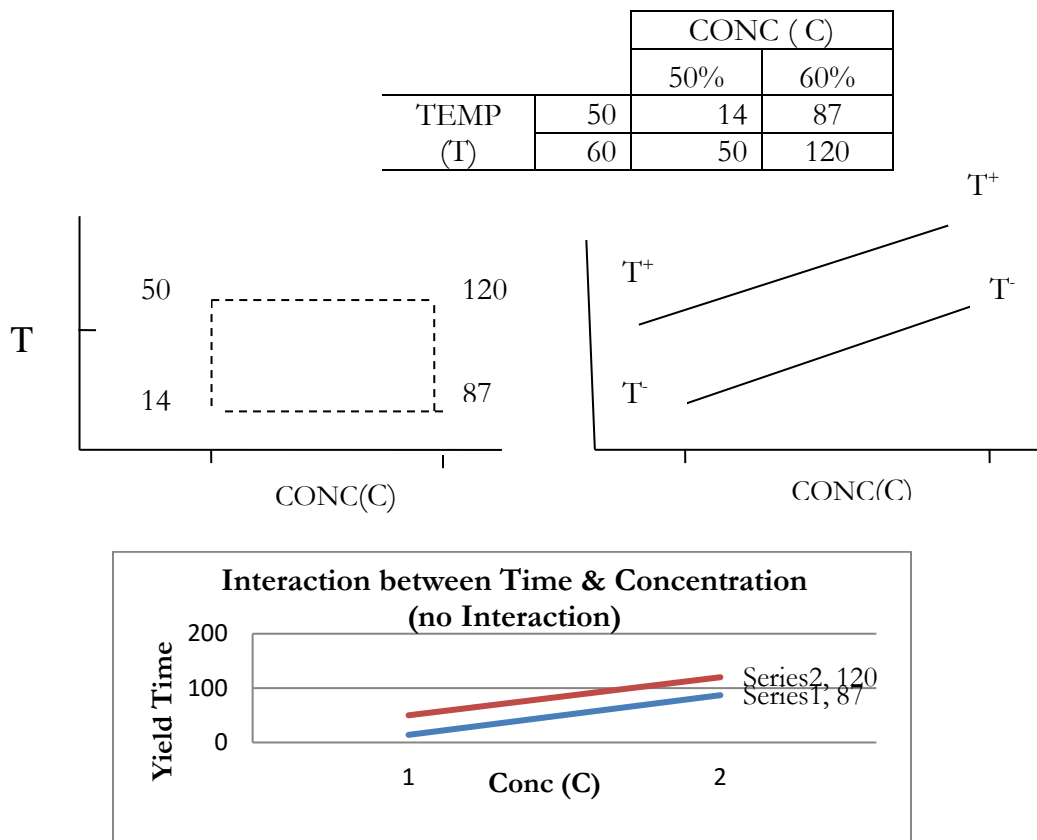


Figure 2 No Visible Interaction Effect



The Design and Analysis of Engineering Experiments II
A SunCam online continuing education course

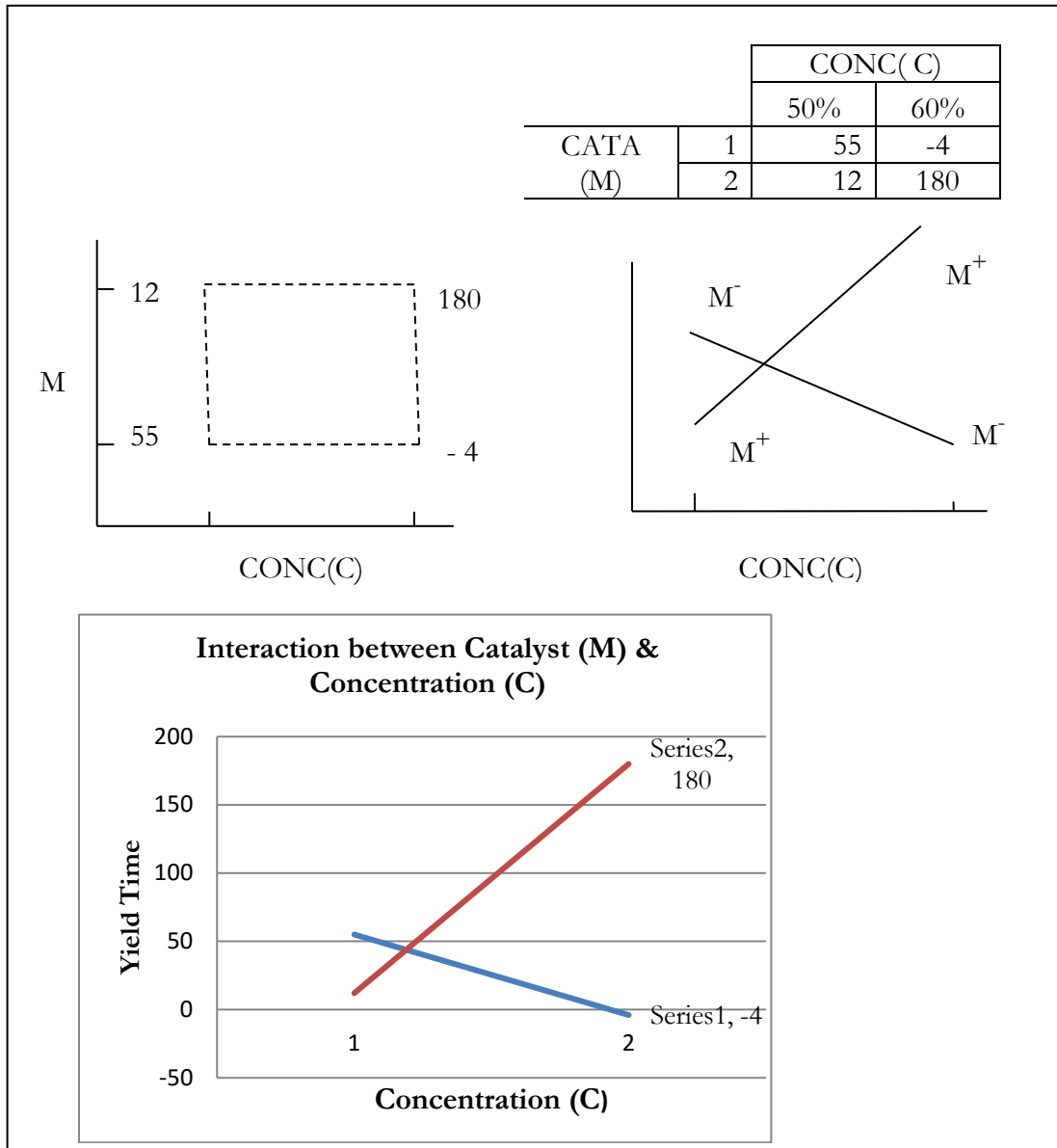


Figure 3 Visible Interaction Effect

Confounding

In factorial experiments, there are practical cases in which randomization is restricted. As an example, restriction can happen in factorial experiment when all data points cannot all be run in one shift or one day. This is more than a question of economics. For example, in a situation where an environmental chamber is to be used to test the reliability of an electronic component, one chamber may not be enough to run the entire test. In such a case a decision must be made about what



The Design and Analysis of Engineering Experiments II

A SunCam online continuing education course

information will be sacrificed. As we indicated earlier, usually the highest level interaction effects or any interaction whose effect is considered not significant is usually confounded for the sake of obtaining the true functional relationship among the factors and the response. Under no circumstances are main effects confounded with blocks.

Thus, confounding is the process by which unimportant comparisons or interactions are deliberately confused for the purpose of assessing the more important comparisons with greater precision. Confounding is required in factorial experiments in which the number of observations capable of being carried out under strictly comparable conditions is less than the number required for the design. Confounding is carried in blocks so that the block effect is confounded with the effect of the interaction chosen for the confounding scheme. In this case, the sum-of-squares for the effect or interaction confounded is exactly equal to the sum-of-squares of all the blocks among which the confounding took place.

To implement a confounding scheme, a defining contrast is first set up. The defining contrast is an expression that indicates which effects (usually interactions) are to be confounded or confused with block. Main effects are **never** used to confound with blocks.

5.1 Confounding, Interaction and the ANOVA Table

When we have confounding, not all treatments can appear in one block because of the restriction imposed on complete randomization. Therefore, some interactions are usually confounded with blocks. When the same interaction is used for all the blocks, similar to what we have so far, then we have complete confounding. In some cases, we may have replication of the experiment, and rather than use the same interaction, we use a different interaction for each replication. When this happens, we have **partial** confounding.

In many situations, the engineer cannot afford several replications of an experiment. Additionally, there is the restriction on randomization which means not all the treatments can be run at one time. This means that confounding will be used, which results in the experiment being run in blocks. As was discussed earlier, the best way to obtain an independent estimate of error is through replication. When there is no replication, some of the higher order interaction terms (order than the ones used for confounding) must be pooled and used as experimental error. The rule of thumb is that when a second order interaction (e.g. ABC) is used as the defining contrast, then all second order interaction and higher are pooled to form the error sum-of-squares. This does not work when a specific interaction is of interest in which it is not used.

There is often some confusion about how interactions are named or labeled. First order interaction involves two factors. A second order interaction involves three factors, while third order interaction involves four factors and so on.

5.2 Methods for Confounding

1. Table of Signs (This is good for 2^n designs)
2. ABD System (This is good for 2^n designs)
3. The Kempthorne method (This is good for 2^n and 3^n designs)



The Design and Analysis of Engineering Experiments II
A SunCam online continuing education course

5.2.1 Table of Signs

Consider three factor A, B, C, assuming no replication, and we are to confound the highest order interaction ABC. In such a case, the defining contrast is: (I, ABC). For the defining contrast, those treatment combinations with the same sign (negative or positive relative to the defining contrast) are assigned to the same block. Whichever block contains '(1)' is the principal block. For the table of signs, write down each treatment combination in the usual order and then write down applicable main effects in the columns (see table 22) as column headings. Also, write down the term or interaction to be confounded as a column heading. To begin, if a main effect appears in the treatment combinations then we assign a positive (+ sign), if not we assign a negative (-sign). Note that for all main effects, the sign with respect to '(1)' is negative. Finally, to obtain the sign for the contrast or the interaction being confounded, multiply the signs of the main effect that make up the interaction or contrast.

Note that this is a simple case of confounding in which the treatment combination is split into two blocks in which one interaction (usually the highest) order is confounded. Higher degrees of confounding involve larger number of blocks. For two blocks only one interaction is needed. For 4 blocks (the blocks in a 2^n must be in multiples of 2) any two interactions appropriately determined will do. Usually the number of block is given by 2^j where j is the number of defining contrast specified. So, if $j=1$, then we have 2^1 which is 2 blocks, if $j=2$, then we have four blocks and so on.

Table 22 table of Signs

Treat. Comb	A	B	C	ABC
(1)	-	-	-	-
a	+	-	-	+
b	-	+	-	+
ab	+	+	-	-
c	-	-	+	+
ac	+	-	+	-
bc	-	+	+	-
abc	+	+	+	+
The Two Block are as follows:				
Principal Block # I		Block # II		
(1)		a		
ab		b		
ac		c		
bc		abc		



The Design and Analysis of Engineering Experiments II
A SunCam online continuing education course

5.2.2 ABD Method (Good for 2^f)

Consider the treatment combinations that we examined earlier. Let the defining contrast be given by {I, ABC}. The procedure to determine what goes into each block is as follows. Those treatment combinations with even number or zero numbers in common with the contrast go in one block. Those with odd number go in another block.

For {I=ABC}:

- (1) Has 0 letters in common with ABC.
- a: has one letter in common with ABC.
- b: has one letters in common with ABC.
- c: has one letter in common with ABC.
- ab: has 2 letters or 0 letters in common ABC (2 modulo 2 =0).
- ac: has 2 letters or 0 letters in common with ABC.
- bc: has 2 letters or 0 letters in common with ABC.
- abc: has 3 letters or 1 letter in common with ABC (3 Mod 2 =1)

Block I (principal block): [(1), ab, ac, bc]; Block II: [a, b, c, abc]

Suppose the defining contrast {I, AB}

For { I=AB}

- '(1)' has 0 letters, 'c' has 0 letters, 'ab' has 2 or 0 letters, 'abc' has 2 or 0 letters
- ' a' has 1 letter; 'b' has 1 letter; 'ac' has 1; 'bc' has 1.

Block I (principal block): [(1), c, ab, abc]; Block II: [a, b, ac, bc]

5.2.3 Kempthorne (Good for 2^f and 3^f Designs)

For the Kempthorne method, we use the defining contrast to specify the defining equation. Each defining contrast in the system of confounding scheme results in a defining equation. The defining equation in general is given as:

$$L = A_1 X_1 + A_2 X_2 + \dots + A_n X_n$$

Where: A_i = the exponent on the ith factor for each defining contrast

X_i = level of ith factor in the treatment combination

Example: Let the defining contrasts for a design be given as {I, ABC, AB²C}

$$L_1 = X_1 + X_2 + X_3$$

$$L_2 = X_1 + 2X_2 + X_3$$

Example: We have a 2³ design with confounding where the defining contrast for the confounding scheme is given as {I, ABC}. Find the elements of the blocks using the Kempthorne scheme. Note that for 2^f design, the processes are modulo 2.

Solution: L= X₁ +X₂ +X₃



The Design and Analysis of Engineering Experiments II
 A SunCam online continuing education course

Table 23 Kempthorne Table with Principal Block

Treat. Comb.	Kempthorne Equation values: (L)
(1)	L=0
a	L=1 (L=1+0+0=1)
b	L=1 (L=0+1+0=1)
ab	L=0 (L=1+1+0=2=0 mod 2)
c	L=1 (L=0+0+1=1)
ac	L=0 (L=1+0+1=2 mod 2 =0)
bc	L=0 (L=0+1+1=2 mod 2=0)
abc	L=1 L=1+1+1=3 mod 2 =1)
Note: those elements with L=0 go in one block and the ones with L=1 go in the other.	
Principal Block	Other Block
(1)	a
ab	b
ac	c
bc	abc

5.3 Confounding and the Elements of Blocks

The principal block is a unique block in the confounding scheme. The following properties of the principal block are useful in determining elements of the principal block. Note that the Principal block is the block that contains the element '(1)'. **Note that in this system 2 mod 2=0**

Property 1. The elements of the Principal form a group closed to multiplication. Thus, when any two or more of the elements or treatment combinations of the principal block are multiplied and the indices appropriately reduced, the resulting product is also an element of the principal block

Property 2. Multiplying the elements of the principal block by any element or treatment combination not in the principal block, generates another block in the system.

Example: From the last example, the principal block (PB) and the other block were as the follows:

Table 24 Continuation of Table 23

PB(I)	Other Block (II)
(1)	a
ab	b
ac	c
bc	abc

Property 1. $(ab) * (1) = ab$
 $(ab) * (ab) = a^2 b^2 = a^0 b^0 = 1 = (1)$
 $(ab) * (ac) = a^2 bc = a^0 bc = bc$
 $(ab) * (bc) = ab^2 c$



The Design and Analysis of Engineering Experiments II
A SunCam online continuing education course

- Property 2.** $(a)*(a) = a^2 = 1 = (1)$
 $(a)*(b) = ab$
 $(a)*(c) = ac$
 $(a)*(abc) = a^2bc = bc$

5.3.1 Generating the Blocks and the elements of the Block

Suppose we have 2^4 and we want to confound in 4 blocks. Since there are 16 total elements, each block will have four elements. We know the number of blocks = 2^j . In this case $j=2$. This means that we have two defining contrasts or two defining equations. The third defining contrast we get by multiplication mod 2. With the two defining equations, we can generate the elements of the four blocks. Note that the number of iterations needed to get the block is also $j=2$.

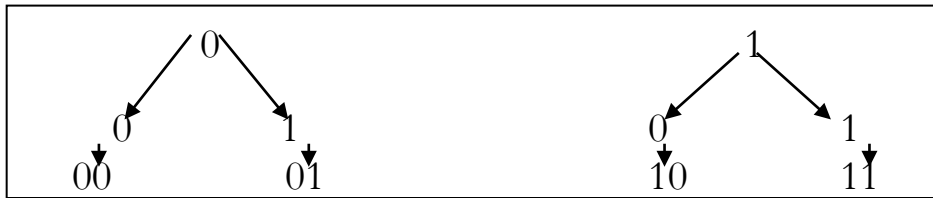
Generate the blocks (or the block headers).

Assume that the defining contrast is given by: {I, ACD, BCD}. "I" typically represents the principal block. We get the 4th block by multiplying the two contrasts mod 2, so that

$$(ACD)*(BCD) = ABC^2D^2 = AB$$

The full set of defining contrast then is: {I, ACD, BCD, AB}

Note that in a 2^f design, the levels are 0,1.



With these we generate the labels: 00, 01, 10, 11

Using I=ACD, BCD

$$L_1 = X_1 + X_3 + X_4, \quad L_2 = X_2 + X_3 + X_4$$

- Write down all the elements including their combinations in this 2^4 design:

$$[(1), a, b, ab, c, ac, bc, abc, d, ad, bd, abd, cd, acd, bcd, abcd]$$

Substitute (1) in both equations, we get 00 \Rightarrow (1) belongs in the first (Principal block).

Substitute 'a', we get 10 \Rightarrow 'a' belongs in the second block.

Substitute 'b', we get 01 \Rightarrow 'b' belongs in the third block.

Substitute 'c', we get 11 \Rightarrow 'c' belongs in the fourth block.

- By inspection, we can quickly get the elements of the principal block. Once we have that we can use the elements that we already determined for the remaining blocks to multiply the elements of the principal block to fill up the remaining blocks.

The remaining elements of the principal blocks are



The Design and Analysis of Engineering Experiments II
A SunCam online continuing education course

'cd' will give $L_1=0, L_2=0 \Rightarrow 00$

'abc' will give $L_1=0, L_2=0 \Rightarrow 00$

'abd' will give $L_1=0, L_2=0 \Rightarrow 00$

We know 'a' is in second block, so multiply 'a' with the elements of the principal block one after the other to get the elements of the block containing 'a', that is block 2

$\Rightarrow (a)^*(1) = \mathbf{a}; (a)^*(cd) = \mathbf{acd}; (a)^*(abc) = a^2bc = \mathbf{bc}; (a)^*(abd) = a^2bd = \mathbf{bd}$

Similarly for b and so on

$\Rightarrow (b)^*(1) = \mathbf{b}; (b)^*(cd) = \mathbf{bcd}; (b)^*(abc) = ab^2c = \mathbf{ac}; (b)^*(abd) = ab^2d = \mathbf{ad}$

00	01	10	11
(1)	a	b	c
cd	acd	bcd	d
abc	bc	ac	ad
abd	bd	ad	abcd

Example. The following data from a 2^3 design with defining contrast $\{I, AC\}$ was run in two blocks. Use the Kempthorne approach to determine the elements of the blocks. Use the Yates scheme to compute the SS for the effects and their interaction and show that, indeed, the effect AC was confounded with blocks.

The model is $y_{ijk} = \mu + A_i + B_j + AB_{ij} + C_k + AC_{ik} + BC_{jk} + ABC_{ijk} + \epsilon_{ijk}$

	A1		A2	
	B1	B2	B1	B2
C1	5 (1)	4 b	0 a	2 ab
C2	-3 c	-1 bc	0 ac	-2 abc

$I=AC, L=X_1+X_3$. Since this is a 2^f design, the blocks are labeled $L=0$, and $L=1$

L=0	L=1
(1)=5	a=0
ac =0	c=-3
b =4	ab=2
abc=-2	bc=-1
+7	-2



The Design and Analysis of Engineering Experiments II
 A SunCam online continuing education course

Treat Comb	Resp.	I	II	III	Divisor r 2 ^f =8	SS=(Col III) ² /Divisor	
(1)	5	5	11	5	8	3.125	
a	0	6	-6	-5	8	3.125	
b	4	-3	-7	1	8	0.125	
ab	2	-3	2	-1	8	0.125	
c	-3	-5	1	-17	8	36.125	
ac	0	-2	0	9	8	10.125	Same as block
bc	-1	3	3	-1	8	0.125	
abc	-2	-1	-4	-7	8	6.125	

$$SS_{Block} = \frac{(7)^2 + (-2)^2}{4} - \frac{(5)^2}{8} = \left(\frac{49+4}{4} \right) - \frac{(25)}{8} = 13.25 - 3.125 = 10.125$$

Since we used AC for confounding, all other first order interaction or higher are added to the error sum-of-squares, that is, (AB, BC, ABC).

Source	df	SS	MS	F	Sig
A	1	3.125	3.125		
B	1	0.125	0.125		
C	1	36.125	36.125		
AC or Block	1	10.125	10.125		
Error (AB, BC, ABC)	3	6.375	2.125		
Total	8	55.875			

Example: Consider a 2⁵ design with confounding in four (4) blocks of 8 treatments each. Since we are not given the defining contrast, we can set up our own. Note that the number of blocks is given by 2^j, where j is the number of defining equations. With 2 blocks, j=2, so we need two defining equations. Let the defining equations be {I, ADE, ABCD}. Since we have four (4) blocks, there are 3 degrees of freedom to account for. We can get the remaining contrast since we need 4 blocks, that is, the principal blocks (I in the defining contrast) plus three more. We can get the remaining contrast by the multiplication of the set of contrasts we have on hand modulo 2, that is : (ADE) * (ABCD) = BCE. So, the full set of defining contrast is given by: {I, BCE, ADE, ABCD}. We choose BCE and ADE for the Kempthorne equation:

$$L_1 = X_2 + X_3 + X_5 \Rightarrow BCE$$

$$L_2 = X_1 + X_4 + X_5 \Rightarrow ADE$$



The Design and Analysis of Engineering Experiments II
 A SunCam online continuing education course

Table 27 Layout of Block Confounding {ADE,ABCD}

(0,0)	(1,0)	(0,1)	(1,1)
(1)	b	a	e
ad	abd	d	ade
bc	c	abc	bce
abe	ae	be	ab
abcd	acd	bcd	abcde
bde	de	abde	bd
ace	abce	ce	ac
cde	bcde	acde	cd

5.4 Confounding in 3^f

Consider a 3³ experiment in which the experiment cannot be completely randomized. Suppose nine can be randomized and run in one shift, nine on another shift and the last nine on the third shift. We have 3³ design in 3 blocks of 9 each. The number of blocks is given by 3^j. In this case, j=1 since 3¹=3. If we choose to confound the interaction, ABC, then the defining contrast is:

{I, ABC}. This requires only two degrees of freedom to be confounded with blocks (3 block requires 2 df) Why does the ABC interaction have eight (8) degrees of freedom? Well, since each factor has three levels, the degree of freedom for each factor is 2. Thus, for ABC interaction, we have (2)(2)(2)=8. So, since ABC interaction has 8 degrees of freedom, we can partition ABC into its four component parts each with 2 degrees of freedom. We can then use one of the four to confound with the blocks. $ABC(8\text{ df}) = [ABC(df = 2), AB^2C(df = 2), ABC^2(df = 2), AB^2C^2(df = 2)]$

Using Kempthorne Equation, we have:

$$L=X_1+X_2+X_3. \text{ The block labels a 0, 1, 2}$$

Table 28 Results of Kempthorne for {I,ABC}

0		1		2	
000	(1)	001	C _L	002	C _Q
111	A _L B _L C _L	112	A _L B _L C _Q	110	A _L B _L
201	A _Q C _L	202	A _Q B _Q	200	A _Q
210	A _Q B _L	211	A _Q B _L C _L	212	A _Q B _L C _Q
102	A _L B _Q	100	A _L	101	A _L C _L
120	A _L B _Q	121	A _L B _Q C _L	122	A _L B _Q C _Q
021	B _Q C _L	022	B _Q C _Q	020	B _Q
012	B _L C _Q	010	B _L	011	B _L C _L
222	A _Q B _Q C _Q	220	A _Q B _Q	221	A _Q B _Q C _L



The Design and Analysis of Engineering Experiments II
A SunCam online continuing education course

Source	df
A	2
B	2
AB	4
C	2
AC	4
BC	4
Block or ABC	2
Error: or AB^2C, ABC^2, AB^2B^2	6 (no replication)
TOTAL	26

Fractional Factorial Design

The idea of a confounding scheme is because there is no opportunity for complete randomization. For example, the shift is too short to accommodate all the runs required to complete the experiment. When that happens, the experiment is run partially and completed at a later time. The total number of treatment combinations to be run in one shift depends on the type of design.

For 2^f design, the treatments per block are in tuples as are the blocks. For example, we can have 2 treatments per block, 8 treatments per block, 16, or 32 treatments per block and so on. Also, in the case of 2^f , we can have 2 blocks, 4 blocks, 6 blocks and so on all in multiples of 2.

For 3^f designs, the treatments per block are in triples as are the number of blocks in the system. Consequently, we have 3 treatments per block, 9 treatments per block, 27 treatments per block all in multiples of 3. The blocks are also in triples, namely, 3 blocks, 9 blocks, 27 blocks, etc.

In fractional factorial designs, while we still use confounding to determine which treatment combinations (interactions) are confused with the block, only a fraction of the experiment can be run. In fractional design, it is not possible to completely randomize or to run the total experimental replications due to size and, hence, leading to the high cost of running the experiment. Thus, in the case of 2^f , we talk about running one-half replicate, one-quarter replicate or one-eighth replicate of the experiment using a defining contrast that would ensure that the important treatment combinations in the design are run. The same with 3^f design. A third-replicate or one-ninth replicate and so on are run. That is why it is important in fractional factorial designs that the confounding scheme and the resulting defining contrast are well thought out so that interactions that are really important are not confounded. Regardless, main effects are never confounded with blocks.



The Design and Analysis of Engineering Experiments II
A SunCam online continuing education course

6.1 Aliases

In fractional designs, we have the issue of aliases. Aliases happen when two or more effects have the same numerical values when confounding is used. Aliases are indistinguishable in terms of their numerical values. Thus, in fractional design, it is important to ensure that they are not both present in the fraction of the experiments that is to be run if the design is to be of any value. What we want to avoid especially is for a main effect to be aliased with another main effect because we cannot extract the significance of the effects. We determine an alias as follows.

Determination of Aliases for 2^f

For each treatment effect and interaction that is not part of the defining contrast, multiply the treatment or interaction with the defining contrast (modulo 2 or modulo 3 depending on the design). The result after the multiplication is the alias. For the sake of an example, suppose the defining contrast was (I=BCE, ADE, ABCD) for a 2⁵ and a one-quarter replicate was run. This means that these interactions are confounded with the four blocks. The aliases for the main effect and some interaction are as follows:

$$A(BCE)= ABCE, A(ADE) =A^2DE=DE, A(ABCD)=A^2BCD=BCD$$

$$B(BCE)=B^2CE=CE, B(ADE)=ABDE, B(ABCD)=AB^2CD=ACD$$

$$C(BCE)=BC^2E=BE, C(ADE)=ACDE, C(ABCD)=ABC^2D=ABD$$

$$D(BCE)=BCDE, D(ADE)=AD^2E=AE, D(ABCD)=ABCD^2=ABC$$

$$E(BCE)=BCE^2=BC, E(ADE)=ADE^2, E(ABCD)=ABCDE$$

$$AB(BCE)=AB^2CE=ACE, AB(ADE)=A^2BDE=BDE, AB(ABCD)=A^2B^2CD=CD$$

It is obvious that this is not a bad design because the main effects are not aliased with any main effects or the fourth-order interaction, that is, the highest order interaction in this case. The highest order interaction in any design with no replication, is pooled to form the SS error .

Determination of Aliases for 3^f

The Alias for 3^f is computed a little differently. The exponent on the first effect in a 3^f is never greater than unity so when it is greater than one, the combination is squared to reduce the exponent modulo 3. For 3^f, multiply the effect by the contrast modulo 3 and again multiply the square of the contrast modulo 3 resulting in two separate alias value per effect. For example for a 3³ design with I=AB²C. The aliases will be as follows: For A: A(AB²C)=A²B²C=(A²B²C)²=ABC² (exponent on first effect can't be greater than 1). Again for A:A(AB²C)²=A(A²B⁴C²)=A³B⁴C²=B⁴C²=BC²; B: B(AB²C)=AC, B(AB²C)²=A²B⁵C²=A⁴B¹⁰C⁴=ABC, etc.



The Design and Analysis of Engineering Experiments II
 A SunCam online continuing education course

6.2 Fractional Factorial for a 2^f

There are two different scenarios in any fractional factorial design. In one scenario, the defining contrast is given and the issue is to determine the treatment combination that will result from the defining contrast. In this case, we can employ the Kempthorne approach or the table of sign to determine the resulting treatment that would apply.

A more difficult situation is where the engineers know the treatment combinations that are required in the experiment and so the question is what is the defining contrast that would yield the given treatment combinations. This is a more difficult problem and would require more detailed effort than we can devote at this time. However, it is an important consideration and much more practical than the other case where the defining contrast is given.

Example: Suppose we have a 2⁵ design where, due to the physical limitation of the system and the cost of experiment, all we can afford is a one-quarter replicate. The defining contrast is not known but we have determined that it is : {I, ACD). The eight data points (1/4 of 2⁴) are as follows

Table 30 Data for the one-quarter replicate of 2⁵ Fractional Design

Treat. Comb	a	ab	c	bc	d	bd	acd	abcd
Response	96.6	125.7	14.1	9.5	43.5	22.4	28.2	52.5

		A1		A2	
		B1	B2	B1	B2
D1	C1			96.6	125.7
	C2	14.1	9.5		
D2	C1	43.5	22.4		
	C2			28.2	52.5

If we use the table of sign, we can setup the sign for the treatment and the interaction in the usual way. Namely, if the treatment exists in the effect shown then we indicate that existence by a positive sign, if not we indicate by a negative sign. The reason is that if the defining contrast did indeed result in those treatment combinations, the sign of each element with respect to the defining contrast will be positive. Note that in a fractional factorial design, the block to be used is chosen at random. For example, the treatment 'a' will have + sign under the effect A negative for all other effects. However, if you multiply the signs for 'a' for the effects that make up the defining contrast (ACD), the sign is positive. This is the same for all the treatments. The aliases with the main effects are:

**Main Effect Aliases: A(ACD)=CD, B(ACD)=ABCD,
 C(ACD)=AD, D(ACD)=AC**

Table 31 Table of Sign to confirm the defining contrast for 2⁵ fractional Design



The Design and Analysis of Engineering Experiments II
A SunCam online continuing education course

Treat Comb.	A	B	C	D	ACD
a	+	-	-	-	+
ab	+	+	-	-	+
c	-	-	+	-	+
bc	-	+	+	-	+
d	-	-	-	+	+
bd	-	+	-	+	+
acd	+	-	+	+	+
abcd	+	+	+	+	+

We will use the YATES scheme to compute the SS. We will indicate the full compliments of a 2^5 in the YATES table but with holes because some values were not actually run. We will proceed the usual way on the YATES table, only now where the values are nonexistent, we will assume they are zeros.

Table 32 YATES Scheme for 2^5 Fractional Factorial Design

Treat Comb	Response	I	II	III	IV	SS= (Col IV) ² /8
(1)		96.6	222.3	245.9	392.5	-
a	96.6	125.7	23.6	146.6	213.5	5671.125
b		14.1	65.9	198.7	27.7	95.91
ab	125.7	9.5	80.7	14.8	79.1	781.10
c	14.1	43.5	222.3	24.5	-183.9	4227.40
ac		22.4	-23.6	3.2	-99.3	1232.56
bc	9.5	28.2	-65.9	33.7	11.7	17.11
abc		52.5	80.7	45.4	-21.3	56.71
d	43.5	96.6	29.1	-198.7	-99.3	Same as AC
ad		125.7	-4.6	14.8	-183.3	Same as C
bd	22.4	-14.1	-21.1	-245.9	-21.3	Same as ABC
abd		-9.5	24.3	146.6	11.7	Same as BC
cd		-43.5	29.1	-33.7	213.5	Same as A
acd	28.2	-22.4	4.6	45.4	392.5	SS Block
bcd		28.2	21.1	-24.5	79.1	Same as AB
abcd	52.5	52.5	24.3	3.2	27.7	Same as B



The Design and Analysis of Engineering Experiments II
 A SunCam online continuing education course

Table 33 ANOVA for 2⁵ Fractional Factorial Design

Source	df	SS	MS	F	F(0.05,1,3) =10.1
A(or CD)	1	5671.125	5671.125	19.9	Sig
B(or ABCD)	1	95.91	95.91	< 1	n.s
C(or AD)	1	4227.40	4227.40	14.83	Sig
D(or AC)	1	1232.56	1232.56	4.33	ns
AB(or BCD)	1	781.1			
BC(ABD)	1	17.11			
BD(or ABC)	1	56.71			
Error (AB+BC+BD)	3	854.92	854.92/3= 284.97		
Total	7				

Please note that we cannot test for the significance of a block because there are no degrees of freedom left since there is only one block. At any rate, the block (ACD) is aliased with the correction factor CF so there is no need to explore this further. You will also notice that B is aliased with the highest order interaction. This is quite unfortunate because the highest order interaction is usually pooled with other non-relevant interaction to form the error sum-of-squares.

6.3 Fractional Factorial for a 3^f

We will demonstrate this design with an example.

The torque (lb) after pre-heat was to be measured on some rubber material with factor A (Temp) at (145°C, 155°C, 165°C), factor B (Mix) at (I, II, III), and factor C (Lab; 1,2,3). The 33 27 experiments seemed too costly, so a 1/3 (one-third) replicate was run with AB²C used as the defining contrast.

Using Kempthorne: L= X₁+2X₂+X₃. The three blocks from the confounding scheme are:

Table 34 Blocks for 3⁴ Fractional Design with AB²C

L=0	L=1	L=2
000	100	200
110	210	010
011	111	211
121	221	021
212	012	112
022	122	222
220	020	120
201	001	101
102	202	002



The Design and Analysis of Engineering Experiments II
 A SunCam online continuing education course

With this design which consists of 3^3 factorial in 3 blocks of nine treatments per block, the effects can be broken down into 13, 2 df effects, namely: A, B, C, AB, AB^2 , AC, AC^2 , BC, BC^2 , ABC, AB^2C , ABC^2 , AB^2C^2 . Now, if only one of these blocks are run (1/3 replicate), say block L=1, the numerical results are shown but now only 8 degrees of freedom since we have only nine data points.

Table 35 Data for 1/3 Replicate of a 3^3 Fractional Design with AB^2C as Defining Contrast

		A1(Temp)			A2(Temp)			A3(Temp)		
		145°C			155°C			165°C		
		B1(Mix)			B2(Mix)			B3(Mix)		
		I	II	III	I	II	III	I	II	III
C1(Lab)	1			16.8	11.2				9.9	
C2(Lab)	2	15.8				14.4				17.8
C3(Lab)	3		17.1				20.5	15.7		

The aliases are:

For A: $A(AB^2C) = A^2B^2C = (A^2B^2C)^2 = ABC^2$, again for A: $A(AB^2C)^2 = A(A^2B^4C^2) = A^3B^4C^2 = BC^2$

For B: $B(AB^2C) = AB^3C = AC$, again for B; $B(AB^2C)^2 = A^2B^5C^2 = A^4B^{10}C^4 = ABC$

For C: $C(AB^2C) = AB^2C^2$, again for C: $C(AB^2C)^2 = A^2B^4C^3 = A^2B^4 = A^4B^8 = AB^2$

For AB: $AB(AB^2C) = A^2B^3C = A^4B^6C^2 = AC^2$, again for AB: $AB(AB^2C)^2 = A^3B^5C^2 = B^2C^2 = B^{10}C^4 = BC$

So, we have the following:

A (ABC^2, BC^2)

B (AC, ABC)

C (AB^2C^2, AB^2)

AB(AC^2, BC)

We can compute the SS using the traditional method:

$$SS_A = \frac{(49.7)^2 + (46.1)^2 + (43.4)^2}{3} - \frac{(139.2)^2}{9} = 6.66, SS_B = \frac{(42.7)^2 + (41.4)^2 + (55.1)^2}{3} - \frac{(139.2)^2}{9} = 38.13$$

$$SS_C = \frac{(37.9)^2 + (48)^2 + (53.3)^2}{3} - \frac{(139.2)^2}{9} = 40.81$$

$$SS_{TOTAL} = (15.8)^2 + (17.1)^2 + (16.8)^2 + (11.2)^2 + (14.4)^2 + (20.5)^2 + (15.7)^2 + (9.9)^2 + (17.8)^2 - \frac{(139.2)^2}{9}$$

$$= 2238.68 - 2152.96 = 85.72$$

$$SS_{AB} = SS_{TOTAL} - SS_A - SS_B - SS_C = 0.12$$



The Design and Analysis of Engineering Experiments II
A SunCam online continuing education course

Table 36 ANOVA for 1/3 Replicate of a 3⁴ Fractional Design with AB²C as Contrast

Source	df	SS	MS	F	F(0.05,1,3)=10.1
A(ABC ² , BC ²)	2	6.66	3.33	55.5	Sig
B(AC, ABC)	2	38.13	19.07	317.83	Sig
C(AB ² C ² , AB ²)	2	40.81	20.41	10.21	Sig
AB(AC ² , BC)	2	0.12	0.06		
Total	8	85.72			

Random, Fixed Effect Model and Expected Mean Square.

In the planning stages of an experiment, the engineer must decide whether the levels of the factors to be run are to be set at fixed levels or chosen at random from many possible levels. A major consideration affecting this choice may be stated as follows. Are the results from such an experiment going to be judged for these levels only or are they to be extended to more levels of which those in the experiment are but a random sample?

As an example, in the case of temperatures, time or pressure, it is usually desirable to pick fixed levels since not all possible levels are practical in a given experiment.

However, in the case of operators, days, batches, or shifts, it is desirable to pick random levels, since the performance of any operator, or a day's production is not as important as whether the operator, batches, or days, in general, increase or decrease the variability of the process in question. It is, of course, ridiculous that one would want to decide whether levels should be fixed or random after the experiment has been performed. When all levels are fixed, we have a fixed model. When all levels are random, we have a random model. However, when some are fixed and some are random, then we have a mixed effects model.

Realistically, most models are mixed because even when all the factors have fixed levels, the error term is always considered random, independent and identically distributed. The question then is how to determine the mean square for the terms in the ANOVA table to properly test for their significance. So far, we have assumed that all the levels are random and so we directly divide the mean square of each term by the mean square error to determine their significance. However, by using the correct mean square or what we now call the expected mean square (EMS) for each term in the model, we can now properly test for their significance.

7.1 Single Factor Model

Let $y_{ij} = \mu + \tau_j + \varepsilon_{ij}$

μ =fixed constant, ε_{ij} ~Normal Independent and Identically Distributed with mean zero, and variance equal to σ_e^2 , ε_{ij} ~NIID (0, σ_e^2).



The Design and Analysis of Engineering Experiments II
 A SunCam online continuing education course

1. If the treatment levels τ_i are fixed, then the ANOVA will look as follows.

Source	df	EMS
τ_i	$k-1$	$\sigma_e^2 + \frac{n \sum \tau_i^2}{k-1}$ ($\sigma_e^2 + n < \sigma_\tau^2 >$)

With respect to the hypothesis test, we will test the hypothesis that the mean effect is zero versus the alternative that some pairs of means are not equal zero. The variance component for the treatment mean is called pseudo-variance because it does not meet the ordinary definition of variance. Note that, in general, the mean square of any term in the model reflects the notion that the variance of any term consists of the variance due to error plus a component due to the term itself.

In this case, our hypothesis is formulated as: $H_0: \tau_i=0, H_1: \tau_i \neq 0$ for some j

2. If the treatment levels τ_i are random, then the ANOVA will look as follows:

Source	df	EMS
τ_i	$k-1$	$\sigma_e^2 + n\sigma_\tau^2$
ε_{ij}	$k(n-1)$	σ_e^2

The hypothesis in this case is as follows: $H_0: \sigma_\tau^2=0, H_1: \sigma_\tau^2 > 0$

7.2 Two-Factor Model

We will demonstrate how we determine the expected means square (EMS) for a given model using a two-factor mixed model. Let $y_{ijk} = \mu + A_i + B_j + AB_{ij} + \varepsilon_{k(ij)}$

where: $i=1,2,\dots,a$ (a levels of factor A), A is fixed
 $j=1,2,\dots,b$ (b levels of factor B), B is random, $k=1,2,\dots,n$ (replication)
 if A is fixed and B is random, then AB is random

The hypotheses to be tested are as follows:

$H_0: A_i=0, \sigma_B^2=0, \sigma_{AB}^2=0$. The EMS table will look as follows

Source	df	EMS
A_i	$a-1$	$\sigma_e^2 + n\sigma_{AB}^2 + nb < \sigma_A^2 >$
B_j	$b-1$	$\sigma_e^2 + na\sigma_B^2$
AB_{ij}	$(a-1)(b-1)$	$\sigma_e^2 + n\sigma_{AB}^2$
$\varepsilon_{k(ij)}$	$ab(n-1)$	σ_e^2

We will now develop a general procedure for populating an EMS column to determine how the significance test will be run for a given model.

7.3 EMS Rules for Establishing an EMS column for ANOVA

Example of a two-factor mixed model: $y_{ijk} = \mu + A_i + B_j + AB_{ij} + \varepsilon_{k(ij)}$

1. Write the factors as row headers in a row/column table

A_i	_____
B_j	_____
AB_{ij}	_____

2. Write down the $\varepsilon_{k(ij)}$ _____ subscripts for each factor as



The Design and Analysis of Engineering Experiments II
A SunCam online continuing education course

column heading. Above each subscript indicate whether the factor for each subscript is fixed or random by writing F or R above each. For the error term, always write R. Above the second row, write the number of levels or replications corresponding to the factors or terms.

	a	b	n
	F	R	R
	i	j	k
A_i			
B_j			
AB_{ij}			
$e_{k(ij)}$			

3. For each row (each factor or term in the model) write down the level or observation on top of each subscript, so long as the subscript does not appear in the row. For example, for A, copy down 'b' under subscript j and n under subscript k, but another under i because i appears in the A row .

	a	b	n
	F	R	R
	i	j	k
A_i		b	n
B_j	a		n
AB_{ij}			n
$e_{k(ij)}$			

4. For any bracketed subscript in the model, place a one under those subscripts which are inside the bracket.

	a	b	n
	F	R	R
	i	j	k
A_i		b	n
B_j	a		n
AB_{ij}			n
$e_{k(ij)}$	1	1	1

5. Fill the remaining cells with 0 if the subscript is for fixed effect (F) and 1 if it is for random (R) effect.

	a	b	n
	F	R	R
	i	j	k
A_i	0	b	n
B_j	a	1	n
Ab_{ij}	0	1	n
$e_{k(ij)}$	1	1	1

6. To find the EMS, for any term in the model, do the following:



The Design and Analysis of Engineering Experiments II
A SunCam online continuing education course

i) Cover in the column or columns which contain un-bracketed subscript letters in this term. For A cover i, for B, cover j for AB cover ij, for ϵ cover k only.

ii). Multiply the remaining number in each row. Each of these products is the coefficient of its corresponding variance term in the model so long as the subscript on the term is also a subscript of the term is being sought. The sum of the coefficients multiplied by their corresponding variances is the EMS of the term under consideration. For example, for A, cover column i. The products of the remaining elements are: bxn for first row, 1xn, for second row, 1xn for third row, and 1x1 for the last row. The product n from the second row is not used because it does not contain i, which is part of the subscript for A that is being considered. So, the resultant EMS for A is:

$$bn\sigma_A^2 + n\sigma_{AB}^2 + \sigma_e^2$$

	a	b	n	
	F	R	R	
	i	j	k	EMS
A _i	0	b	n	$\sigma_e^2 + n\sigma_{AB}^2 + nb < \sigma_A^2 >$
B _j	a	1	n	$\sigma_e^2 + na\sigma_B^2$
AB _{ij}	0	1	n	$\sigma_e^2 + n\sigma_{AB}^2$
e _{k(ij)}	1	1	1	σ_e^2

7.3.1 Testing for Significance: A vs AB, B vs e_{k(ij)}, AB vs e_{k(ij)}

7.4 Rule for Determining the Degrees of Freedom for EMS

Factor	Subscript	Rule	df
A _i	i	i-1	a-1
B _j	j	j-1	b-1
AB _{ij}	ij	(i-1)(j-1)	(a-1)(b-1)
AB _{j(i)}	j(i)	(j-1)i	a(b-1)
e _{k(ij)}	k(ij)	(k-1)ij	(n-1)ab

Nested or Hierarchical Designs

In the design of experiments, it is not always the case that all the factors are crossed with each other the result of which captures our traditional definition of a factorial experiment. From practical considerations, it is not always possible to cross levels of one factor with the levels of another factor. We bring up this design because it is perhaps use of the most common designs that engineers face and the one most commonly misused or misunderstood. Thus, it is important to provide the tool necessary to handle it. Consider a situation where we want to take the readings of a compression test on a particular specimen. There are four machining centers available for this test and each machining center has four different heads. At first thought, we might consider this as factorial design with two factors machines (M) and heads (H) where each machine is randomly crossed with each head and then a typical factorial analysis is carried out with test of significance for the factors and their interaction. The model that describes this usual factorial design model is:



The Design and Analysis of Engineering Experiments II
 A SunCam online continuing education course

$$Y_{ijk} = \mu + M_i + H_j + MH_{ij} + \epsilon_{k(ij)}, \text{ assuming } k \text{ replications.}$$

The question that persists in this type of thinking is the following. Was every machine really crossed with all heads? The answer of course is no because that would require removing the heads from each machine and refitting them to all other machines, which may be possible but not probable. Since a set of heads are tied only to set machines and only those, then the model has to reflect this relationship because the heads are nested within each machine and not crossed with all the machines. The recognition of this fact leads us to a new modeling scheme where we reflect the fact that the heads are nested within the machine as opposed to being stand alone and crossed among all machines. The revised model will now look as follows:

$Y_{ijk} = \mu + M_i + H_{j(i)} + \epsilon_{k(ij)}$, and we will analyze this model with the EMS approach. The EMS table determines how the test of significance is carried out. This is unlike the regular ANOVA table where the test was based on the ratio of the mean square of each term error means square.

Example: Assume that in the scenario described we have Compression tests on tungsten specimens were run on three machines each with three heads. The data is as follows:

Table 37 Data Layout for Nested/Hierarchical Design

Mach	M1				M2				M3				
Head	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	
	1.5	1.5	2.7	3	1.9	2.3	1.8	1.9	2.5	3.2	1.4	7.8	
	1.7	1.6	1.9	2.4	1.5	2.4	2.9	3.5	2.9	5.5	1.5	5.2	
	1.6	1.7	2	2.6	2.1	2.4	4.7	2.8	3.3	7.1	3.4	5	
Head Total	4.8	4.8	6.6	8	5.5	7.1	9.4	8.2	8.7	15.8	6.3	18	
Mach Total	24.2				30.2				48.8				103.2

We assume for this design that we have a random sample of heads that can be used on a given machine., then we have a hierarchical or nested design in which the machines are fixed and the heads random. $Y_{ijk} = \mu + M_i + H_{j(i)} + \epsilon_{k(ij)}$, $i=1..3$, $j=1,..4$, $k=1,..3$

	3	4	3	
	F	R	R	
	i	j	k	EMS
M_i	0	4	3	$\sigma_e^2 + 3\sigma_H^2 + 12\langle \sigma_M^2 \rangle$
$H_{j(i)}$	1	1	3	$\sigma_e^2 + 3\sigma_H^2$
$\epsilon_{k(ij)}$	1	1	1	σ_e^2



The Design and Analysis of Engineering Experiments II
 A SunCam online continuing education course

$$SS_{Total} = (1.5)^2 + (1.5)^2 + \dots + (3.4)^2 + (5.0)^2 - \frac{(103.2)^2}{36} = 381.44 - 295.84 = 85.60$$

$$SS_{Mach} = \frac{(24.2)^2 + (30.2)^2 + (48.8)^2}{12} - CF = 323.26 - 295.84 = 27.42, df = 2$$

$$SS_{H(M)} = SS_{H(M_1)} + SS_{H(M_2)} + SS_{H(M_3)}$$

$$SS_{H(M_1)} = \frac{(4.8)^2 + (4.8)^2 + (6.6)^2 + (8.0)^2}{3} - \frac{(24.2)^2}{12} = 51.213 - 48.803 = 2.41$$

$$SS_{H(M_2)} = \frac{(5.5)^2 + (7.1)^2 + (9.4)^2 + (8.2)^2}{3} - \frac{(30.2)^2}{12} = 78.753 - 76.00 = 2.753$$

$$SS_{H(M_3)} = \frac{(8.7)^2 + (15.8)^2 + (6.3)^2 + (18.0)^2}{3} - \frac{(48.8)^2}{12} = 229.673 - 198.453 = 31.22$$

$$SS_{H(M)} = SS_{H(M_1)} + SS_{H(M_2)} + SS_{H(M_3)} = 2.41 + 2.753 + 31.22 = 36.383, df = (3)(3) = 9$$

$$SS_{Error} = SS_{Total} - SS_M - SS_{H(M)} = 85.60 - 27.42 - 36.383 = 21.80, df = (2)(3)(4) = 24$$

Table 38 ANOVA for the Machine Heads EMS Example

Source	SS	df	MS	EMS	F
M_i	27.42	$i-1=2$	13.71	$\sigma_e^2 + 3\sigma_H^2 + 12\sigma_M^2 >$	13.71/4.04
$H_{j(i)}$	38.38	$(j-1)i=9$	4.04	$\sigma_e^2 + 3\sigma_H^2$	4.04/0.91
$e_{k(ij)}$	21.8	$(k-1)ij=24$	0.91	σ_e^2	
Total	85.6	35			

8.1 Testing for significance on an EMS Table

- M vs H, $F=3.93$ ($F_{0.01, 2, 9}=8.06$); therefore not significant
- H versus Error, $F=4.42$ ($F_{0.01, 9, 24}=2.3$); heads effect is significant from machine to machine.

Note that not all the terms or factors/interactions are tested for significance with the error mean square as has heretofore been the case.

Regression Analysis

The purpose of Regression Analysis is not to explain what type of data should or should not be collected for any given purpose. The purpose is to explain some of the techniques used in extracting information, such as the main features of the relationship between the variables in the data, and to make a strong case for the importance of properly designing the experiment to collect needed data.

In any system where quantities change, it is of interest to look at the effects, if any, of some variables or others. Indeed, there may be a functional relationship which may be approximated by a



The Design and Analysis of Engineering Experiments II
A SunCam online continuing education course

simple mathematical relationship. In other instances, the functional relationship may be complicated. Still, there are situations where no meaningful relationship seems to exist, but we might wish to express or relate them by some sort of mathematical equations.

There are two main types of variables, namely predictor or independent variables (X), and the response or dependent variables (Y).

A regression equation is a prediction equation fitted to a set of experimental data values to describe a possible relationship between a single dependent variable Y and one or more independent variables X. In the case of a single Y and a single X, the situation becomes a regression of Y on X. For n independent variables, it becomes the regression of Y on X_1, X_2, \dots, X_n .

One particular method commonly used in expressing the relationship between the variables is the method of least squares. In this method, the unknown parameters are estimated under certain assumptions and a fitted equation is obtained. The value of the equation can be examined by substituting known values to see its predictability.

The method of least squares in regression analysis will be used to examine the data and to draw conclusions about any functional relationship between the response and the independent variables.

The simplest kind of regression is the bi-variate linear regression.

$$\text{Model: } Y = f(x), \quad \text{i.e., } Y = \alpha + \beta x$$

which can better be expressed as: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, where $\varepsilon_i \sim \text{NIID}(0, \sigma_e^2)$

This equation can be expanded from the bi-variate to a polynomial regression. Under the least squares method, we can also solve the multivariate or the multivariable linear regression which is expressed as:

$$Y = f(X_1, X_2, \dots, X_n) = A_0 + A_1 X_1 + A_2 X_2 + \dots + A_n X_n$$

We can also have a Multivariate polynomial regression. For a polynomial of 2nd degree, we will have the following Model: $Y = A_0 + A_1 X_1 + A_2 X_2 + A_{11} X_1^2 + A_{22} X_2^2 + A_{12} X_1 X_2$

Nonlinear Regression can be handled with the least squares method if linear transformation is possible.

9.1 Model Solution.

Let a model be represented by: $Y = \beta_0 + \beta_1 x_i + \varepsilon_i$,

where ε_i = error (sometimes called the residual) and has zero mean and a given distribution.

Let $f(x) = b_0 + b_1 X_i$ be the predicted i^{th} y value (when $X = X_i$) and

$$b_0 = \text{estimate of } \beta_0, \quad b_1 = \text{estimate of } \beta_1$$

Then the deviation of the observed value of y from the line y is: $\varepsilon_i = y_i - f(x_i)$

The aim of the curve fitting effort is to minimize this deviation. Specifically, the aim is to minimize the SS of the error (deviation). The procedure used to accomplish this is the method of least squares. Define Q as:

$$Q = \sum e_i^2 = \sum (Y - f(x))^2 = \sum [Y_i - (b_0 + b_1 X_i)]^2$$

A way to solve the model is to develop normal equations arising from the model. A way to do that is to set up an expression for the least squares estimate and then optimize by taking partial derivatives



The Design and Analysis of Engineering Experiments II
 A SunCam online continuing education course

with respect to the parameters of the model. Since only two parameters (β_0, β_1) are involved, we take partials of Q with respect to those two parameters and optimize.

$$\frac{\partial Q}{\partial a} = -2 \sum_{i=1}^n (y_i - b_0 - b_1 x_i)$$

$$\frac{\partial Q}{\partial b} = -2 \sum_{i=1}^n (y_i - b_0 - b_1 x_i)(x_i)$$

$$nb_0 + b_1 \sum x_i = \sum y_i \dots\dots\dots(1)$$

$$b_0 \sum x_i + b_1 \sum x_i^2 = \sum x_i y_i \dots\dots\dots(2)$$

$$b_0 = \sum y_i - b_1 \sum x_i \Rightarrow b_0 = \frac{\sum y_i}{n} - b_1 \frac{\sum x_i}{n} \Rightarrow b_0 = \hat{\beta}_0 = \bar{y} - b_1 \bar{x}$$

$$b_1 = \hat{\beta}_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

Let : $S_{xy} = n \sum xy - \sum x \sum y, \quad S_{xx} = n \sum x^2 - (\sum x)^2, b_1 = \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$

Equations one and two are called normal equations. The dimensions of the set of equations (no of these normal equation is equal to the number of parameters) are such that the system of equations is solvable so long as the data is not ill due to very significant differences in the magnitude of the data. In other words, in a matrix form the matrix always has an inverse so long as the determinant is not zero.

As the model becomes larger, due to the increase in the number of parameters in the model, solving for the estimates becomes more tedious and time consuming. In such a case, we will resort to matrix algebra. One of the advantages of a matrix approach is that once the problem has been formulated in matrix form, the solution can be applied to any size problem.

9.2 Matrix Approach

Let: $Y = \beta_0 + \beta_1 X_i + \epsilon_i \Rightarrow Y = X\beta + \epsilon$

Define: $Y =$ Vector of observations from the experiment-- (n x1 vector)

$X =$ Matrix of independent variables-- (nx2) Matrix

$X^T =$ A transpose of the X matrix

$\beta =$ Vector of parameters to be estimated-- (2x1) vector

$\epsilon =$ Vector of errors or deviations (nx1) vector

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \bullet \\ \bullet \\ y_n \end{bmatrix}; \quad X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \bullet & \bullet \\ \bullet & \bullet \\ 1 & x_n \end{bmatrix}; \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}; \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \bullet \\ \bullet \\ \epsilon_n \end{bmatrix}, \quad X^T \begin{bmatrix} 1 & 1 & \bullet & \bullet & 1 \\ x_1 & x_2 & \bullet & \bullet & x_n \end{bmatrix}$$



The Design and Analysis of Engineering Experiments II
A SunCam online continuing education course

$$X^T X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdot & \cdot & 1 \\ x_1 & x_2 & \cdot & \cdot & x_n \end{bmatrix} = \begin{bmatrix} n & \sum x \\ \sum x & \sum x^2 \end{bmatrix}, \quad X^T Y = \begin{bmatrix} 1 & 1 & \cdot & \cdot & 1 \\ x_1 & x_2 & \cdot & \cdot & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} \sum y \\ \sum xy \end{bmatrix}$$

The normal equations can be rewritten in matrix form as:

$$(X^T X)\hat{\beta} = X^T Y \\ \Rightarrow \hat{\beta} = (X^T X)^{-1}(X^T Y),$$

where $\hat{\beta}$ is the vector of the estimates of the parameters.

$$(X^T X)\hat{\beta} = X^T Y \Rightarrow \begin{bmatrix} n & \sum x \\ \sum x & \sum xy \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \sum y \\ \sum xy \end{bmatrix} \\ \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} n & \sum x \\ \sum x & \sum xy \end{bmatrix}^{-1} \begin{bmatrix} \sum y \\ \sum xy \end{bmatrix}$$

Table 39 Data for the Regression Model									
i	1	2	3	4	5	6	7	8	9
X_i	1.5	1.8	2.4	3.0	3.5	3.9	4.4	4.8	5.0
Y_i	4.8	5.7	7.0	8.3	10.9	12.4	13.1	13.6	15.3

Example: Given the data of table 39, use the method of least squares to determine the parameters of the model. $Y=b_0+b_1X$

$$n = 9, \sum x_i = 30.3, \sum y_i = 91.1, \sum x_i y_i = 345.09, \sum x_i^2 = 115.11, \sum y_i^2 = 1036.65$$

$$CF = \frac{(\sum y)^2}{9} = 922.134, \bar{X} = 3.3667, \bar{Y} = 10.122$$

$$S_{XY} = \frac{\sum x_i y_i - (\sum x_i \sum y_i) / n}{n} \\ = \frac{345.09 - ((30.3)(91.1) / 9)}{9} = 38.387$$

$$S_{XX} = \frac{\sum x_i^2 - ((\sum x_i)^2 / n)}{n} \\ = \frac{115.11 - ((30.3)^2 / 9)}{9} = 13.1$$

$$b_1 = \hat{\beta}_1 = \frac{S_{XY}}{S_{XX}} = \frac{38.387}{13.10} = 2.9303,$$

$$b_0 = \hat{\beta}_0 = \bar{Y} - b_1 \bar{X} = 10.22 - (2.9303)(3.3667) = 0.2545$$

$$S_{yy} = \sum (y_i - \bar{y})^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{n} = 1036.65 - 922.134 = 114.516$$



The Design and Analysis of Engineering Experiments II
 A SunCam online continuing education course

$$X^T X = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} = \begin{bmatrix} 9 & 30.3 \\ 30.3 & 115.11 \end{bmatrix} \Rightarrow \begin{bmatrix} 9 & 30.3 \\ 30.3 & 115.11 \end{bmatrix}^{-1} = \frac{1}{117.9} \begin{bmatrix} 115.11 & -30.3 \\ -30.3 & 9 \end{bmatrix}$$

$$X^T Y \begin{bmatrix} 91.1 \\ 345.09 \end{bmatrix}, \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \frac{1}{117.9} \begin{bmatrix} 115.11 & -30.3 \\ -30.3 & 9 \end{bmatrix} \begin{bmatrix} 91.1 \\ 345.09 \end{bmatrix} = \begin{bmatrix} 0.2545 \\ 2.95 \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

Source	df	df	SS	SS (Matrix Form)	SS	MS	F	F
b ₀	1	1	$[(\sum y)^2 / n] = CF$	$Y^T 11^T Y = CF = 922.134$	-	-	-	
b ₁ b ₀	1	1	$b_1 S_{XY}$	$\beta^T X^T Y - CF$	112.485	112.5	26.00	
error	n-2	7	Subtraction	$(Y^T Y - \beta^T X^T Y)$	2.04	0.291		
Total	n-1	8	$\sum y_i^2 - CF$	$Y^T Y - CF$	114.52			

$$\beta^T (X^T Y) = [0.2568 \quad 2.95] \begin{bmatrix} 91.1 \\ 345.09 \end{bmatrix} = 23.394 + 1011.217 = 1034.611, Y^T Y = \sum y_i^2 = 1036.65$$

$$SS_{Error} = Y^T Y - B^T X^T Y = 1036.65 - 1034.611 = 2.04$$

$$Y^T 11^T Y = [y_1 \quad \cdot \quad \cdot \quad y_n] \begin{bmatrix} 1 \\ \cdot \\ \cdot \\ 1 \end{bmatrix} x [1 \quad \cdot \quad \cdot \quad 1] \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = (1 \times n \times n \times 1) x (1 \times n \times n \times 1) = a \text{ scalar}$$

9.3 Coefficient of Determination (R²)

The coefficient of determination R² is a measure of the fit of the regression line. It represents the amount or proportion of variation in the data that is explained by the fitted model or regression line. It measures how well the model fits the data. The higher this value the better the model. In our example the fit of the model or model accuracy is about over 98% which is very good.

$$R^2 = \frac{SS \text{ Regression}}{\text{Total Corrected Sum of squares}} = \frac{b_1 S_{xy}}{S_{yy}} = \frac{2.9303(38.39)}{114.52} = 0.9823$$

9.4 A Note about the Least Squares Method

The X^TX (called the X transpose X) matrix is a symmetric square matrix. Because it is a square matrix, we can always find an inverse of the matrix except when the determinant is zero or close to zero. This happens when the magnitude of the differences in the data is very high. A way to overcome this problem is to use transformation, such as the log or square root transformation or appropriate scaling. We did not discuss hypothesis testing for the estimates, but those can be accomplished by properly extracting the variance estimates for the parameters.



The Design and Analysis of Engineering Experiments II

A SunCam online continuing education course

Summary

This second course in the series covers some practical aspects of the design and analysis of engineering experiments. There are no theoretical developments and most of the computations use basic arithmetic. Several important and practical issues are addressed including factorial designs.

The issue of missing values due to a tool break or momentary loss of power, for example, has been addressed with numerical examples and calculations. Missing values is a different issue from incomplete designs. Incomplete designs happen when, in the intended design, only certain number of days or shifts for example are available to run the experiment.

The issue of system physical limitations in handling the entire design was addressed using confounding schemes where certain (unwanted or unimportant) interactions are confounded with blocks in order to assess more important effects and/or interactions. Fractional designs have been addressed as a way to handle limited resources where it is not practical or cost effective to run complete experiments.

Nesting or hierarchical designs were introduced to address the practical situation where complete randomization or crossing of factors is not possible. Regression using the method of least squares was introduced as way to estimate the parameters of the proposed model. The matrix approach was advanced as part of the least squares approach when higher order models are required.

REFERENCES

1. Taguchi, G., Chowdhury, S., Taguchi, S., (2000), Robust Engineering: Learn How to Boost Quality While Reducing Costs & Time to Market 1st Ed, McGraw Hill, New York, NY
2. Taguchi, G., Elsayed, A. E., Hsiang, T.C., (1989), Quality Engineering in Production Systems, McGraw- Hill Series in Industrial Engineering and Management Science, NY
3. Montgomery, D.C., and Runger, G. (2011), Applied Statistics and Probability for Engineers, 5th ed., John Wiley and Sons, NY.
4. Walpole, R., Meyers, R.H. (2002), Probability and Statistics for Engineers and Scientists, 7th ed., Prentice- Hall Inc., Upper Saddle River, NJ.
5. Devore, J.L., (2012), Probability and Statistics for Engineering and the Sciences, 8th ed., Duxbury & Brooks/Cole, Boston, MA.
6. Montgomery, D. (2012), Design & Analysis of Experiments, 8th ed., Wiley & Sons, NY.
7. Hinkelmann, K., Kempthorne, O., (1994), Design and analysis of experiments, Volume I, John Wiley and Sons, NY
8. Petersen, R. G., (1985), Design and analysis of experiments, Marcel Dekker, New York, NY
9. Hicks, C, R., Turner, K. V. (1999), Fundamental Concepts in the Design of Experiments, 5th Ed, Oxford University Press, New York, NY
10. Hicks, C., R. (1963), Fundamental Concepts in the Design of Experiments, 5th Ed, Holt, Rinehart and Winston, New York, NY