Design of Experiments (DOE or DOX) WITH SEVERAL FACTORS

Reference : CHAPTER 11 of Devore's 7th Edition **Maghsoodloo**

An experiment involving 2 or more factors is said to be a (full or complete) factorial design iff observations are taken at every possible factor level combination (FLC) of all the factors involved. A factorial experiment is balanced iff the size of the sample at each FLC (or in each cell) is the same, namely n. If sample sizes for different cells differ, the design is said to be unbalanced. As an example, consider an injection molding process where the output is the strength of the molded piece measured in psi, and the 2 inputs that may have an effect on strength are Temperature (A) and Pressure (B). Factor A is 2 levels (100, 150 ^{0}F), and factor B is at 3 levels (50, 75, 100 psi). This is called a 2×3 factorial design and clearly it has 6 = 2×3 FLCs as illustrated in Tables 5 & 6. In Table 5, $y_{11.} = 79$, $y_{12.} = 55$, $y_{13.} = 133$, $y_{21.} = 82$, $y_{22.} = 113$, and $y_{23.} = 132$.

Table 5. (An unbalanced Factorial Design)

A \ B	50 psi	75	100
100 ^{o}F	34, 45	55	48, 40, 45
150	42, 40	55, 58	50, 42, 40

Note that there are 6 cells or FLCs.

An Unbalanced 2×3 Factorial Experiment

Sample Sizes: $n_{11} = 2$, $n_{12} = 1$, $n_{13} = 3$,..., $n_{23} = 3$.

The design in Table 5 is CR (completely randomized), i.e., the order of taking all the 13 observations was completely randomized. Table 6 below provides a balanced factorial with n = 3 observations per cell. Again the experiment in Table 6 is a 2×3 CRD (CR design) factorial, where both factors are quantitative and their levels are fixed (i.e., not randomly selected). As in the case of the one-factor experiment, USS = $36^2 + 28^2 +...+ 50^2 = 37{,}912$ (with 18df); CF = $810^2/18 = 36{,}450$ (with 1 df) → SS_T = USS – CF =1462.00 (with 17 df). The error sum of squares is computed as SS(Error) = $SS_{11} + SS_{12} + ... + SS_{23} = (36^2 + 28^2 + 33^2 - 97^2/3) + (55^2 + 60^2 + 59^2 - 174^2/3) + ... + (52^2 + 53^2 + 50^2 - 155^2/3) = 32.66667 + 14.0 +$

48.66667 +12.66667 + 128.0 + 4.66667 = 240.66667 (with 2×6 =12 df) → SS (Model) = SS_T – SS_{Error} = 1221.33333 (with 5 df).

Table 6. (A balanced factorial with n = 3 responses, or replicates, per cell)

B A	50 psi	75	100	$y_{i..}$
100 F^o	36, 28, 33 (97)	55, 60, 59 (174)	47, 39, 38 (124)	395
150	38, 41, 43 (122)	54, 46, 38 (138)	52, 53, 50 (155)	415
$y_{.j.}$	219	312	279	$y_{...}$= 810

It is paramount to understand that all the calculations that we have performed thus far apply to every factorial design whether balanced or not. As shown above, the model does have 5 df, 1 of which is absorbed by factor A, 2 df are carried by B, and there remains 2 df for what is called the interaction effect between factors A and B denoted by A×B, i.e.,

$$SS(model) = SS(A) + SS(B) + SS(A\times B) \qquad (52)$$

Note that the df on the LHS of equation (52) must "jive" with those on the RHS of equation (52), i.e., 5 = 1 + 2 + 2, and therefore, we have an orthogonal decomposition of SS(Model) into 3 components due to the fact that the design is balanced. As a result,

$$SS(A) = \frac{395^2 + 415^2}{9} - \frac{810^2}{18} = 22.2222\bar{2},$$

$$SS(B) = \frac{219^2 + 312^2 + 279^2}{6} - \frac{810^2}{18} = 741.00$$

and from Eq. (52), SS(A×B) = SS(Model) – SS(A) – SS(B) = 458.111111 (with 2 df). Again, it is essential to be cognizant of the fact that the above procedure for computing SS of different effects (except for Total, Model and Error) is valid only for Balanced Factorial Experimental Designs. If the design is unbalanced, then the General Linear Model (GLM) procedure must be applied in order to compute SS of different effects in the Model. Unfortunately, the GLM procedure is well outside the scope of this course, except for analyzing unbalanced data using Minitab.

Before we define the interaction between 2 factors in impacting an output Y, we present the balanced design in Table 6 in an orthogonal array (OA) format.

OA

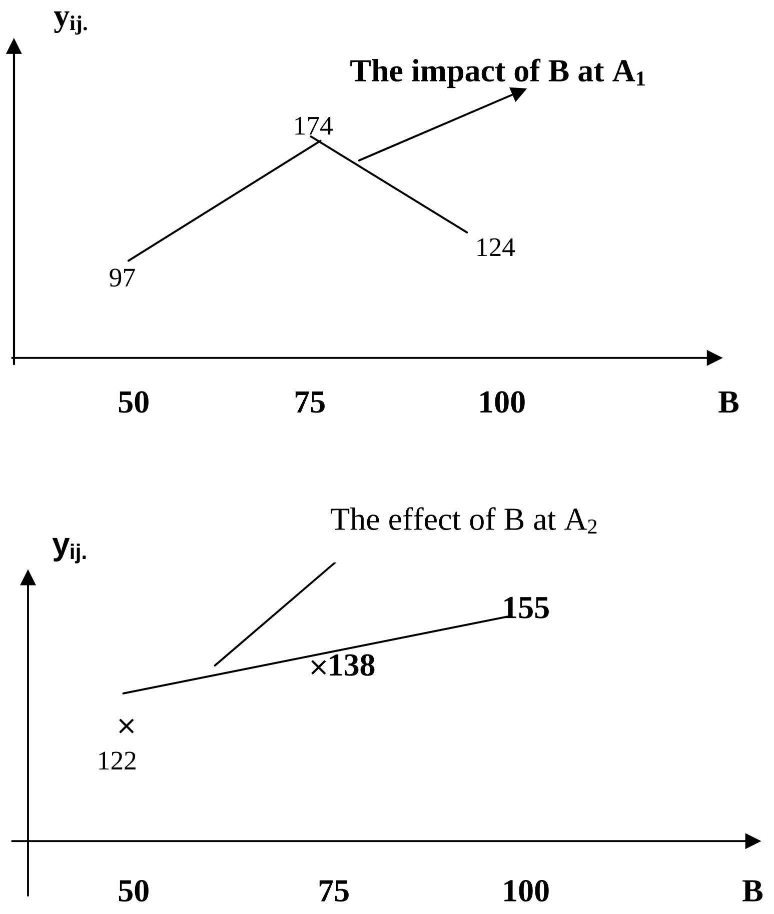
A	B	y_{ijk}	$y_{ij.}$
100	50	36, 28, 33	97
100	75	55, 60, 59	174
100	100	47, 39, 38	124
150	50	38, 41, 43	122
150	75	54, 46, 38	138
150	100	52, 53, 50	155
		$y_{...}$ =	810

Note that the array on the left is orthogonal because every FLC (A_i, B_j) i =1, 2 and j =1, 2, 3 occurs exactly once in the design matrix. Because the factorial design has 6 FLCs, the orthogonal array (OA) has exactly 6 rows. These 6 rows (or FLCs, or cells) with subtotals $y_{ij.}$ carry exactly 5 df amongst them; since the model also has 5 df, then we can also obtain SS(Model) as follows:

$$\text{SS(Model)} = \frac{97^2 + 174^2 + 124^2 + 122^2 + 138^2 + 155^2}{3} - \frac{810^2}{18} = 1221.3333\overline{3}, \text{ as before!}$$

<u>**Exercise 93.**</u> **For the unbalanced design of Table 5, compute the SS(Model) in 2 different ways to determine if both methods yield the same value of SS(Model)! (b) Then compute SS(A), SS(B), and SS(A×B) as you would compute them for a balanced factorial experiment. (c) Use Minitab (proc GLM) to also verify that the SS of different effects in the model cannot be computed in the same manner that are computed for a balanced design.**

<u>**Definition.**</u> **Two factors, A & B, interact as they impact the output Y iff the effect of B at the i^{th} level of A is significantly different from the effect of B at the k^{th} ($k \neq i$) level of A for at least one k. As an example, consider the balanced CRD of Table 6, where the effect of factor B at A_1 is negatively quadratic (i.e., concave downward), and the effect of B at A_2 is almost completely linear as illustrated in Figures 26. Therefore, A and B interact in affecting the response variable Y because the effect of B on Y at different levels of A are dissimilar. The ANOVA is provided in Table 7 at the bottom of the following page. Note that when all factors in the experiment are fixed, then all hypotheses are tested against MS(Error); otherwise, that is not the case. We now develop the identities for breaking**



Figures 26. The interaction effect of factors A and B on Y.

Table 7. (The ANOVA for Table 6 Data)

Source	**df**	**SS**	**MS**	**F_0**	***P-value***
Total	**17**	**1462.00**			
Model (or Between cells)	**5**	**1221.3333**	**244.2666**	**12.1795**	**0.00023**
A	**1**	**22.22222**	**22.22222**	**1.1080**	**0.31324**
B	**2**	**741.00000**	**370.5000**	**18.4737**	**0.00022**
A×B	**2**	**458.11111**	**229.0555**	**11.4211**	**0.00167**
Pure Error (or Within cells)	**12**	**240.66667**	**20.05556**		

down the Total SS of a 2-factor experiment only when the design is balanced.

$$y_{ijk} \equiv \mu + (\mu_{i..} - \mu) + (\mu_{.j.} - \mu) + (\mu_{ij.} - \mu_{i..} - \mu_{.j.} + \mu) + (y_{ijk} - \mu_{ij.}) \qquad \textbf{(53a)}$$

$$y_{ijk} \equiv \mu + [A_i + B_j + A\times B_{ij}] + \in_{ijk} \equiv \mu + [\text{Model effects}] + \in_{ijk} \qquad \textbf{(53b)}$$

The 4th term on the RHS of (53a) represents the interaction effect between A and B, and the last on the RHS gives the error term; further, every term in (53), except for y_{ijk} and $\in_{ijk}$, is a population parameter. As in the case of a single-factor design, we replace every parameter in (53a) with its point sample estimator in order to obtain:

$$y_{ijk} \equiv \bar{y}_{...} + (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}) + (y_{ijk} - \bar{y}_{ij.}) \qquad \textbf{(54)}$$

We now transpose $\bar{y}_{...}$, the 1st term on the RHS of (54), to the LHS .

$$y_{ijk} - \bar{y}_{...} \equiv (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}) + (y_{ijk} - \bar{y}_{ij.}) \qquad \textbf{(55)}$$

$$\textbf{or} \qquad y_{ijk} - \bar{y}_{...} \equiv \hat{A}_i + \hat{B}_j + \widehat{(A\times B)}_{ij} + e_{ijk} \quad .$$

Next, we square both sides of (55) and use the fact that the triple sum of all cross-product terms are zero in order to obtain the orthogonal decomposition of Total SS (sum of squares).

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{...})^2 = nb\sum_{i=1}^{a}(\bar{y}_{i..} - \bar{y}_{...})^2 + na\sum_{j=1}^{b}(\bar{y}_{.j.} - \bar{y}_{...})^2 +$$

$$+ \; n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 + \sum\sum\sum(y_{ijk} - \bar{y}_{ij.})^2 \qquad \textbf{(56)}$$

In Eq. (56), the 1st term on the RHS represents SS(A), the 2nd term is SS(B), the 3rd term is SS(A×B), and the last term on the RHS represents SS(Error) = SS(Residuals). Further, "*a*" represents the number of levels of factor A and "*b*" represents the number of levels factor B. Equation (56) clearly shows that SS(Total) = SS(Model) + SS(Error), where the model terms now are A, B, and A×B. However, only when the design is balanced, do we have an orthogonal partition of SS(Model) as given below.

SS(Model) = SS(A) + SS(B) + SS(A×B).

<u>Exercise 94.</u> **Prove that**

$$SS(A\times B) = n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 \equiv \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij.}^2/n) - CF - SS(A) - SS(B).$$

Exercise 95. Work Exercises 17, 18, and 20 on pages 418 – 419 of Devore's 7th edition. Present the data of exercise 11-18 and 11-20 in OA formats. (b) Show that in the balanced case (i.e., $n_{ij} = n$ for all cells), $V(e_{ijk}) = (n-1)\sigma_{\in}^2/n$. As a result, the studentized residual $r_{ijk} = e_{ijk}/\sqrt{(n-1)MS_{RES}/n}$.

DOE INVOLVING MORE THAN TWO FIXED FACTORS

Consider an experiment where the impact of three inputs "Feed Rate (A)", "Depth of Cut (B)", and "Tool Angle (C)" on the output "y = Surface Roughness" is being studied in a metal cutting operation. Then, there are 3 factors A (Feed Rate), at 2 levels (20 in/min, 30), B (Depth of Cut) at 2 levels (0.025 inch, 0.04) and C (Tool Angle) at 2 levels (15°, 25). Therefore, the design base is 2, and this is called a 2×2×2 or a 2^3 factorial CRD experiment with n = 2 observations per cell; the experimental layout is shown in an OA format in Table 8. For the sake of convenience and familiarizing you with the base-2 design notation, we let "−1 " represent the low level of a factor and "+1" represent the high level of a factor. That is, level "–1" of factor A in actuality will represent 20 inch/minute, and the factor level combination, (FLC), (1, −1, 1) means that the experimenter has set the feed rate at 30 inch/min, the depth of cut at 0.025 inches, and the tool angle at 25 degrees, etc. Further, note that the response variable, y, is an STB (Smaller-The-Better) type QCH.

The OA in Table 8 clearly shows that the design matrix has 8 (= 2^3) rows or FLCs, and hence the model must have 7 df. Factors A, B, and C each have 1 df, the three 2-way interactions A×B, A×C, and B×C also each have 1 df, and thus we have accounted for 6 df so far out of the 7 for the Model. There remains 1 df, which is absorbed by the 3-way (or 2nd-order) interaction A×B×C. Due to the orthogonality of the design, we must have

$$\text{SS(model)} \equiv \text{SS(A)}+\text{SS(B)}+\text{SS(C)}+\text{SS(A}\times\text{B)}+\text{SS(AC)}+\text{SS(BC)}+\text{SS(ABC)} \qquad (57)$$

$$(7)\ \text{df} \equiv (1) + (1) + (1) + (1) + (1) + (1) + (1)\ \text{df}$$

We now proceed to illustrate how to compute different SS using the data in Table 8. As always, SS(Total) = $9^2 + 7^2 + ... + 14^2 - 177^2/16$ = USS – CF = 92.9375 (with 15 df). SS(Error) = $(9^2 + 7^2 - 16^2/2) + ... + (16^2 + 14^2 - 30^2/2) = 2 + 0.5 + 2 + 2 + 2 + 4.5 + 4.5 + 2$ =19.5 (with 8 df) → SS(Model) = SS(Total) – SS(Error) = 73.4375 (with 15 – 8 = 7 df).

Table 8 (A = Feed Rate, B = Depth of Cut, C = Tool Angle, n = 2)

A	B	C	A×B	A×C	B×C	ABC	FLC	y_{ijkr}	$y_{ijk.}$
-	-	-	+	+	+	–	(1)	9, 7	16
-	-	+	+	–	–	+	c	11, 10	21
-	+	-	–	+	–	+	b	9, 11	20
-	+	+	–	–	+	–	bc	10, 8	18
+	-	-	–	–	+	+	a	10, 12	22
+	-	+	–	+	–	–	ac	10, 13	23
+	+	-	+	–	–	–	ab	12, 15	27
+	+	+	+	+	+	+	abc	16, 14	30

We may also compute SS(Model) using the FLC subtotals $y_{ijk.}$:
SS (Model) = $(16^2 + 21^2 + 20^2 + ...+ 30^2) / 2 - CF = 73.4375$ (with 7 df as before). Finally, we compute the SS of different inputs in the model.

$$SS(A) = (75^2 + 102^2) / 8 - CF = 45.5625 \text{ (with 1 df)}$$
$$SS(B) = (82^2 + 95^2) / 8 - CF = 10.5625 \text{ (with 1 df)}$$
$$SS(C) = (85^2 + 92^2) / 8 - CF = 3.0625 \text{ (with 1 df)}$$

The above three SS's clearly show that the impact of factor A is the strongest on the output Y = surface roughness. In order to compute SS(A×B), we cross factors A and B as shown in Table 9, ignoring factor C and within cell variations all together. Table 9 has 4 cells (20, 0.025), (20, 0.040), (30, 0.025), and (30, 0.040) which carry 3 df amongst them. The ANOVA from Minitab is provided on my website. One df is absorbed by A, 1 df by factor B, and 1 by their interaction A×B. Therefore, $(37^2 + 38^2 + 45^2 + 57^2)/4 - CF = SS(A) + SS(B) + SS(A\times B) = 63.6875 \rightarrow SS(A\times B) = 63.6875 - SS(A) - SS(B) = 7.5625$. Similarly, SS(A×C) can be computed using the A×C interaction table; you should verify for your own benefit that SS(A×C) = 0.0625.

Exercise 96. Develop the B×C table and use it to verify that SS (B×C) = 1.5625 for the above 2^3 balanced factorial experiment.

We are now in a position to compute SS(A×B×C). Using Eq. (57), SS(ABC) =

Table 9 **A×B**

B / A	0.025”	0.040	$y_{i...}$
20 in/min	37	38	75
30	45	57	102
$y_{.j..}$	82	95	$y_{....}$ = 177

SS(Model) – SS(AB) – SS(AC) – SS(BC) – SS(A) –SS(B) – SS(C) = 5.0625.

<u>Exercise 97.</u> **(a) Consider the data of Exercise 29 on page 473 of Devore. Give the data layout in the OA format, using A_i for the i^{th} day, B_j for the j^{th} level of temperature, and C_k for the k^{th} screw machine. (b) Obtain the ANOVA table showing all your calculations in neat detail interpreting the F_0 statistics in the ANOVA table. (c) Compute only the 2 studentized residuals pertaining to y_{2232}, and y_{3114}, where the index r runs over the 4 observations within the same cell. Note that for the balanced case the *se*(e_{ijkr}) $= \sqrt{(n-1)\times MS_{RES}/n}$.**

DOE FOR BASE–2 BALANCED FACTORIALS

The notation 2^k is used to denote a factorial experiment involving k factors (A, B, C, D, ..., K) each at 2 levels. For example, the notation 2^4 implies we have 4 factors (A, B, C, D) each at 2 levels (low = –1 or 0, and high = +1 or 1). Since 2^4 = 16, we have 16 FLCs, the corresponding OA will have 16 rows, and the exponent 4 implies we can write 4 columns arbitrarily, one column per factor. Further, there is also a specific notation that is applicable <u>only</u> (and only) to FLCs of base-2 designs (i.e., all factors must be at 2 levels). For the case of 2^4 factorial, the symbol "ac" represents the FLC where factors A and C are at their high levels, while factors B and D are at their low levels. In other words, the presence of a small letter indicates the corresponding factor is at its high level, while the absence of a small letter indicates the factor is at its low level. When all factors are at low levels, then the symbol (1) is used to represent the FLC (–1, –1, –1, –1, ..., –1) = (0, 0, ..., 0). Further, the notation “b” implies that factors A, C, and D are at their low levels while B is

at its high level, etc. The Table 10 illustrates this concept very well. Note that in Table 10, the 4 columns pertaining to the 4 main factors were written arbitrarily, while the interaction columns were obtained by simple multiplication of the factors involved in the interaction. Table 10 shows that the total effect (or contrast) of the 2nd-order interaction ACD is computed as follows: Contrast$_{ACD}$ = – (1) + a – b + ab + c – ac + bc – abc + d – ad + bd - abd – cd + acd – bcd + abcd. Since there are 8 pair-wise comparisons (with n observations per FLC), the average effect of ACD interaction is computed

Table 10. (A 2^4 complete factorial with n = 1 response per cell)

D	C	B	A	FLC	ABCD = A×B×C×D	ACD = A×C×D	y_{ijkL}
-	-	-	-	(1)	+	-	2
-	-	-	+	a	-	+	7
-	-	+	-	b	-	-	8
-	-	+	+	ab	+	+	12
-	+	-	-	c	-	+	5
-	+	-	+	ac	+	-	9
-	+	+	-	bc	+	+	11
-	+	+	+	abc	-	-	14
+	-	-	-	d	-	+	2
+	-	-	+	ad	+	-	4
+	-	+	-	bd	+	+	9
+	-	+	+	abd	-	-	13
+	+	-	-	cd	+	-	15
+	+	-	+	acd	-	+	19
+	+	+	-	bcd	-	-	8
+	+	+	+	abcd	+	+	10

$y_{....}$ = 148

$$\text{from } \overline{ACD} = \frac{\text{Contrast (ACD)}}{8n} = \frac{\text{Contrast}}{n\,2^{k-1}} = \frac{(75-73)^2}{1\times 2^3} = 0.50$$

<u>Exercise 98.</u> Work exercises 44 and 45 on page 441 of your text by obtaining their

ANOVA Tables, and for the time being ignore all references to blocking and the Yates algorithm.

Only in base-2, the contrast for any effect can be written quickly by using the odd-and-even rule. Use the odd-rule for effects with odd number of letters such as A, B, ACD, and ABD, i.e., odds receive a positive sign while evens receive a negative sign, and vice a versa for even effects such as AB, CD, ABCD.

CONFOUNDING IN BLOCKS FOR BASE-2 DESIGNS

It is often impossible to run all the 2^k observations in a 2^k factorial design in a single block, e.g., in a 2^3 design it may be possible to run only 4 experiments in one day, and the other 4 FLCs must be run the next day. In this case, days would form blocks, and the experimenter has to be careful not to confound a main factor, or a 1st-order interaction with blocks. As a second example, four batches of raw material may be needed to conduct 32 experiments in a chemical process; so, batches would form 4 blocks. Referring back to the 1st example, if the experimenter haphazardly runs the 4 FLCs [(1), a, c, ac] on day one, and the FLCs [b, ab, bc, abc] on day 2, then in fact he/she has confounded the effect of factor B with blocks. On the other hand, if the FLCs [(1), a, bc, abc] are run on day 1, and [b, c, ab, ac] are run on day 2, then BC interaction is confounded with blocks. For another example, study pages 434– mid 436 of your text.

To develop a general confounding scheme for base-2 designs, we must use the algebra in base-2, i.e., we must learn that 2 = 4 = 6 = 8 =...= 0 (mod 2), while 3 = 5 = 7 = 9 =... = 1 (mod 2). Recall that the algebra in base 2 has only 2 elements, namely 0 and 1, and this is why 0 is used to represent the low level of a factor, and 1 represents the high level of a factor for base-2 designs. Further, the symbol 2 has no meaning in base 2, and 2 of base 10 is represented by 10 in base 2, i.e., $2_{10} = 10_2$, $3_{10} = 11_2$, etc. Note that each factor has only 2 levels (0, 1), not 3 levels in which case the symbol 2 would be needed to represent the 3rd level if a factor had 3 levels.

For the sake of illustration, consider the experiment of Example 12-11 of Montgomery and Runger (1994, pp. 753-756), where the effects of 4 factors (A = Target Type, B = Seeker Type, C = Target Altitude, D = Target Range) each at 2 levels on the

terminal miss distance (y measured in nearest feet) of a shoulder-fired ground-to-air-missile is being studied. Here we have a 2^4 factorial, as described by the above two authors, but in order to save time the experimenters decided to use 2 operators, which act as 2 blocks. Since one full replicate of a 2^4 factorial provides 16 FLCs, such a design provides a total of 15 df for studying the 15 effects A, B, C, D, AB, AC, AD, BC, BD, CD, ABC, ABD, ACD, BCD, and the 3rd –order interaction A×B×C×D, each having 1 df. However, in this experiment each one of the 2 operators will run 8 of the FLCs, and we must assign the 8 FLCs to each operator in such a manner that none of the lower-order effects are sacrificed, i.e., it is best to sacrifice the effect of the 4-way interaction ABCD. Put differently, the 2 operators carry one df between them and ABCD interaction also has one df, and as a result the assignment of the 8 FLCs to each operator must be carried out in such a manner as to confound the ABCD effect with blocks (or with operators). To start our confounding scheme, we define the contrast function (cf)

$$\xi = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4$$

where x_1 refers to the levels of factor A , x_2 to the levels of B, etc, and α_i = only 0 or 1 in base 2. Note that if our factors had 3 levels, then α_i = 0, 1, or 2. For the Example 12-11 of Montgomery & Runger, the contrast function for the defining contrast (or Generator) ABCD is

$$\xi\,(ABCD) = x_1 + x_2 + x_3 + x_4\,,$$

while the contrast function for sacrificing the ABD interaction is

$$\xi\,(ABD) = x_1 + x_2 + \qquad x_4.$$

Bear in mind that for base-2 designs, the value of the contrast function ξ can be only 0 or 1. We now begin the assignment of the16 FLC's to the 2 blocks, one with $\xi(ABCD) = 0$ and the other with $\xi(ABCD) =1$.

The FLC (1): $\xi(ABCD) = 0 + 0 + 0 + 0 = 0$; the FLC a : $\xi= 1 + 0 + 0 + 0 = 1$; the FLC b : $\xi = 0 + 1+ 0 + 0 =1$; the FLC ab: $\xi = 1 + 1 + 0 + 0 = 2 = 0$ (mod2); the FLC c: $\xi = 1$; the FLC ac: $\xi = 2 = 0$ (mod 2); the FLC bc: $\xi = 2 = 0$ (mod 2); the FLC abc: $\xi= 1 + 1 + 1 + 0 = 3 = 1$ (mod 2), and so on up the FLC abcd for which $\xi(ABCD) = 1 + 1 + 1 + 1 = 4 = 0$ (mod 2). The 2 blocks pertaining to the 2 operators are shown below. Henceforth, the block for which $\xi = 0$ is called the principal block (PB).

The PB →

The Block with ξ = 0: (1) = 3, ab = 7, ac = 6, bc = 8, ad = 10, bd = 4, cd = 8, abcd = 9 feet

The ξ = 1 Block →

The Block with ξ = 1: a = 7, b = 5, c = 6 , abc = 6, d = 4, abd = 12, acd = 9, bcd = 7 feet

<u>Exercise 99.</u> For the above example, use the generator BCD to assign the 16 FLCs to the 2 operators (or blocks) each containing 8 FLCs.

The 16 observations pertaining to the 16 FLCs are given within the above blocks and are repeated here in order of base–2 FLCs ignoring blocks: 3, 7, 5, 7; 6, 6, 8, 6; 4, 10, 4, 12; 8, 9, 7, 9. The ANOVA table was obtained by Excel and provided on my website. To illustrate how the SS's in that spreadsheet were computed, we compute SS (A×D).

$$\mathrm{SS}(A\times D)=\frac{[\mathrm{Contrast(AD)}]^2}{n\sum_{i=1}^{16}c_i^2}=$$

$$=\frac{[(1)-a+b-ab+c-ac+bc-abc-d+ad-bd+abd-cd+acd-bcd+abcd]^2}{1(1^2+(-1)^2+\ldots+1^2)}$$

$$=\frac{13^2}{16}=10.5625.$$

<u>Exercise 99.</u> (b) Verify that SS (BD) = 0.5625 for the above experiment confounded in 2 blocks with g = ABCD as the block generator. (c) Re-compute SS(AD) above by crossing the factors A and D, plus the fact that SS(A) = 27.5625 and SS(D) = 14.0625.

Finally, it is possible that the block size is too small to run even 1/2 of the FLCs within one block, and it is necessary to divide up the 2^k FLCs into 4 blocks, each containing 2^{k-2} runs (generally $k \geq 5$). Since the 4 blocks have 3 df amongst them, then we must confound 3 df with blocks. As an example, consider a 2^5 factorial design run in 4 blocks of 8 FLCs each. Then we need to sacrifice the highest order interactions. This leads to one possible confounding scheme where g_1 = ABC, g_2 = CDE, and $g_3 = g_1 \times g_2$ = ABC^{2}DE =

ABC^0DE (mod 2) = ABDE. The generators g_1 and g_2 are independent while g_3 is not independent of the other 2 generators because $g_3 = g_1 \times g_2$ = ABDE is called the generalized interaction between ABC and CDE. The 4 Blocks have the contrast function values ($\xi_1 = \xi_2 = 0$), ($\xi_1 = 0, \xi_2 = 1$), ($\xi_1 = 1, \xi_2 = 0$), and ($\xi_1 = \xi_2 = 1$).

<u>Exercise 100.</u> (a) Obtain the 4 blocks for the above 2^5 factorial design using the independent generators g_1 = ABC and g_2 = CDE. (b) Work Exercise 45 on page 482 of your text run in 2 blocks confounding ABD with blocks. Use Excel to check your answers.

FRACTIONAL FACTORIALS FOR BASE-2 DESIGNS

As the number of factors in a 2^k factorial increases, the number of experimental runs (or FLCs) required for one complete replicate rapidly outgrows the available resources of most experimenters. One full replicate of a 2^8 factorial requires 256 experimental observations. The problem with such a 2^8 factorial design is 2-fold:

(1) Perhaps too many runs giving rise to much (prohibitive) cost.

(2) Out of the 255 total df, only 8 are absorbed by the main 8 factors, and only ${}_8C_2$ = 28 df are absorbed by the 1st-order interactions AB, AC, AD, ..., GH. The remaining 255 – 36 = 219 df (or 85.88%) correspond to the 3-way, 4-way, ..., 8-way interactions. These high-order interactions (HOIs), if found significant, are difficult to interpret and in some cases are assumed nonexistent. Therefore, sometimes unfortunately in my opinion, their SS's are pooled together and used as the error (or residual) term in the ANOVA table. I would recommend against such pooling!

The above problem worsens more rapidly for the 3^k and 5^k factorials; e.g., one full replicate of a 3^6 factorial requires 729 experiments (or FLCs), where only $(72/728) \times 100$ = 9.89% of the total df are absorbed by effects through the 1st order.

For a 2^k factorial, if it can reasonably be assumed that HOIs are not quite as important, then information on the main factors and 1st-order interactions can be gleaned by running only a FRACTION (i.e., 1/2, 1/4th , 1/8th, etc) of a complete replicate. Such designs are called Fractional Factorials. Fractional factorial designs (FFDs), or fractional replicates, are widely used in industrial research for the prime purpose of fine-tuning a process and improving process and product quality. Process and quality engineers who do not understand how to use FFDs, and therefore do not comprehend the enormous potential

of such designs for continuous quality improvement, are in a terrible disadvantage compared to engineers who do. Manufacturing companies with managers and engineers not knowledgeable about the powers of DOE very rarely survive in global competition and are generally doomed to ruin.

FFD for base-2 can also be used as screening experiments where many factors (generally more than 10) are considered with the purpose of identifying those factors that have large linear effects relative to others. The influential factors, i.e., those with large linear effects on the response Y, are then more completely investigated in subsequent confirmation experiments, either with a full factorial design or with a FFD in base 3 or 5.

THE 1/2 FRACTION OF A 2^K FACTORIAL DESIGN

Consider a 2^4 factorial (4 factors each at 2 levels) and suppose that the experimenter has sufficient resources to conduct only 8 experiments. Because 8 = $\frac{1}{2}(2^4)$, this leads to a 1/2 fraction of a full replicate of a 2^4 factorial, and clearly $(1/2)2^4 = 2^{4-1} = 2^3 = 8$ FLCs. Therefore, we will have 2 fractions (or blocks) each with 8 FLCs. Just like the case of block confounding, we have to be very careful not to lose the effects of the main factors and 1st-order interactions. Further, in order to maximize design resolution (defined later), it is generally best to sacrifice the highest order interactions. Since the 2 fractions (or blocks) carry 1 df between them, we must sacrifice exactly 1 df, namely one of the 5 interactions ABC, ABD, ACD, BCD, or ABCD. If we use one of the four 3-way interactions as the design generator, g, then we have a resolution III design, denoted by 2_{III}^{4-1}. This is due to the fact that any one of the four 2nd-order interactions (ABC, ABD, ACD, BCD) are represented by 3 letters! Hence, we decide to use g = ABCD as the generator in order to attain a resolution IV design (i.e., ABCD has 4 letters)! Note that the resolution of a FFD is simply the minimum number of letters among all generators of the design. The contrast function (cf) for sacrificing g = ABCD is $\xi = x_1 + x_2 + x_3 + x_4$, whose values for base-2 designs can be only 0 or 1. The use of this cf and mod 2 algebra leads to the following 2 fractions (or blocks), each with 8 FLCs with generator g = ABCD.

The PB :	**The Block with $\xi = 0$: (1), ab, ac, bc, ad, bd, cd, abcd**

Note that the fraction for which ξ = 0 is always called the principal block (PB) for all FFDs,

The Fraction with ξ = 1 : a, b, c, abc, d, abd, acd, bcd

and the notation "bd" represents the FLC (0, 1, 0, 1); abc = (1, 1, 1, 0), d = (0, 0, 0, 1), etc.

Exercise 101. **Use the signs under the ABCD column of Table 11, page 174 of these notes, to show that for the above (1/2) 2^4 FFD, the ABCD interaction is also the identity of the design, i.e., *I* = +ABCD for the ξ = 0 block, and *I* = – ABCD is the identity for the ξ=1 block.**

To understand what information is lost when a (1/2) fraction (say the principal block) of a 2^4 factorial is run, consider Table 11 atop the next page , where the 8 FLCs that belong to the principal block: (1), ab, ac, bc, ad, bd, cd, abcd is listed. Recall that –1 refers to low and +1 to high level of a factor. In Table 11, we also have constructed the signs for each of the 15 effects. The response y, the coded by 500, represents etch rate of Silicon Nitride, which is an LTB type QCH. (Sorry that I could not precisely align all the rows in Table 11 unless I used different fonts.) The 4 factors that may impact the response y are A = Anode-cathode gap at 2 levels (0.8, 1.2 cm), B = Pressure in the reactor chamber (2 levels 450 and 550), C = C_2F_6 gas flow (at 125 and 200), and factor D = Power applied to the cathode (at 275 and 325); borrowed from *J. of Solid State Technology*, May 1987, pp127-132.

Table 11 shows that the signs under column A and column BCD are identical, i.e., effects A and BCD are aliased (or confounded together, or indistinguishable from each other); therefore, for this (1/2) fraction A = BCD. Similarly, Table 11 shows that B = ACD, C = ABD, D = ABC, AB = CD, AC = BD, and AD = BC. Further, the effect of ABCD has been sacrificed (or lost completely) because all the signs under the ABCD column are positive, thus the effect of ABCD cannot be studied (because the levels under ABCD do not change). Because the above ½ fractional replicate has 8 FLCs and thus 7 total df, then only 7 effects through 2nd order can be studied which are A, B, C, D, AB = CD, AC = BD, and AD = BC. Another way to obtain the alias structure of a FFD is to multiply an effect by its

Table 11. (The 1/2 fraction of a 2^4 factorial)

The FLCs in the PB	A	B	AB	C	AC	BC	ABC	D	AD	BD	ABD	CD	ACD	BCD	I=ABCD	Response y_{ijkl} Etch Rate LTB
(1)	–	–	+	–	+	+	–	–	+	+	–	+	–	–	+	$50 = y_{1111}$
ab	+	+	+	–	–	–	–	–	–	–	–	+	+	+	+	150
ac	+	–	–	+	+	–	–	–	–	+	+	–	–	+	+	142
bc	–	+	–	+	–	+	–	–	+	–	+	–	+	–	+	101
ad	+	–	–	–	–	+	+	+	+	–	–	–	–	+	+	249
bd	–	+	–	–	+	–	+	+	–	+	–	–	+	–	+	552
cd	–	–	+	+	–	–	+	+	–	–	+	+	–	–	+	575
abcd	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	229

identity elements mod 2 (but use mod 3 for base -3 designs, and mod 5 for base-5 designs). For our example, A = A×*I* = A(ABCD) =A^2BCD = A^0BCD = BCD, and so on!

<u>Exercise 102.</u> (a) Obtain the aliases of the other 12 effects for the above (1/2) fraction in Table 11 using mod 2 algebra. (b) Develop a (1/2) fraction of a 2^4 factorial using ABD as the design generator, i.e., let I = ABD, and then obtain the alias structure of your FFD. Comment on the deficiency of this design!

For the sake of illustration, we compute the contrast value and SS for the effect A in Table 11. Contrast(A) = – (1) + ab + ac – bc + ad – bd – cd + abcd = – 50 + 150 + 142 – 101 + 249 – 552 – 575 + 229 = –508 = Contrast(BCD → SS(A) = SS(BCD) = $(–508)^2$/8 = 32258.

<u>Exercise 102(c).</u> Now compute the contrast values for the other 12 effects in Table 11 and their corresponding SS's. Then use the orthogonality property to verify that the values of your SS's are indeed correct!

THE (1/4)TH FRACTION OF A 2^K FACTORIAL EXPERIMENT (k > 4)

It is often possible that there are too many factors affecting an output of a manufacturing process so that experimenters do not have sufficient resources to conduct even a 1/2 fraction of a full replicate. In such cases resort has to be made to a smaller fraction (or a smaller number of runs) such as (1/4)th , (1/8)th, (1/16)th, etc. However, an experimenter must be cognizant of the fact that for each block of FLCs that he/she does not conduct, he/she is creating exactly one alias for each effect, e.g., a (1/4)th FFD has 3 aliases for each effect because there are 3 blocks of FLCs that are not studied, etc.

Now, consider the Experiment (described in Example 8-4 on pages 298-303 of D. C. Montgomery's 6th edition, 2005) that was performed in an injection-molding process that involved 6 factors each at 2 levels. In order to reduce a part's shrinkage, a team of QC engineers decided to use DOE to identify the controllable factors (or process parameters) that may have large impact the output y = amount of shrinkage. Then the levels of the influential factors would be used to reduce shrinkage (i.e., to improve quality). Factor A = mold temperature, B = screw speed, C = holding time, D = cycle time, E = gate size, and factor F = holding pressure, again each factor is at 2 levels. However, due to lack of resources, the QC engineers could not conduct a complete replicate of the 2^6 factorial which would require $2^6 = 64$ experimental runs (or FLCs) and settled on running only a (1/4)th of the 64 FLCs. In designing any FFD, the first objective must be not to lose the effects of the main factors (A, B, C, D, E, F) and then, if possible, 1st-order interactions (AB, AC, AD, ..., EF), i.e., maximize design resolution such that main factors will not become aliased with 2-way interactions (if the design provides sufficient df). We next describe the procedure for obtaining a $(1/4)^{th}$ fraction of a 2^6 factorial design.

<u>Step 1.</u> Since $(1/4)2^6 = 16$, the 64 FLCs (1), a, b, ab, c, ac, …, bcdef, abcdef must be divided up into 4 blocks each containing 16 FLCs, and as before the fraction that includes the FLC (1) = (–1, –1, –1, –1, –1, –1) is called the principal block (PB).

<u>Step 2.</u> Since the 4 blocks absorb 3 df, we must sacrifice 3 df. Because $(1/4)2^6 = (1/2^2)2^6 = 2^{6-2}$, we need 2 independent generators plus one that is their product Mod 2. Note

that this pattern will persist for all FFDs (base-2, 3, 5, etc., as long as the design base is a prime number), i.e., for a 2^{k-p} FFD we need exactly p independent generators. For example, a 2^{7-3} FFD needs 3 independent generators while a 2^{8-4} needs 4 independent generators. Referring back to the example under consideration with 2 independent generators g_1, g_2 and each generator with 1 df, we note that something is not jiving! Recall that the 4 blocks of 16 FLCs carry 3 df amongst them, and therefore the other df to be sacrificed must be the generalized interaction $g_3 = g_1 \times g_2$.

<u>Step 3.</u> Select the p (= 2 for our design) generators in such a manner as to maximize design resolution (which is the minimum number of letters in all the $2^p - 1$ generators). For our example the QC engineers selected g_1 = ABCE, and g_2 = BCDF. Then our 3rd generator is $g_3 = (ABCE) \times (BCDF) = AB^2C^2DEF = AB^0C^0DEF$ (mod 2) = ADEF. Since the minimum number of letters in all 3 generators is 4, then the design resolution is R = IV.

<u>Step 4.</u> Appropriately separate the 2^k FLCs into 2^p blocks. This can be done the hard way by defining p contrast functions (for our example ξ_1 and ξ_2). Recall that for base-2 designs the values of a contrast function can be only 0 or 1. Therefore, for our example, the 4 blocks will pertain to ($\xi_1 = \xi_2 = 0$, the PB), ($\xi_1 = 0$, $\xi_2 = 1$), ($\xi_1 = 1$, $\xi_2 = 0$), and ($\xi_1 = \xi_2 = 1$). These contrast functions are $\xi_1 = x_1 + x_2 + x_3 + x_5$ for ABCE, and $\xi_2 = x_2 + x_3 + x_4 + x_6$ for BCDF. Next, obtain the PB by computing the values of ξ_1 and ξ_2 for all the 64 FLCs, starting with (1) = (0, 0, 0, 0, 0, 0) and through abcdef = (1, 1, 1, 1, 1, 1). Clearly $\xi_1((1)) = \xi_2((1)) = 0$.

We now determine the cf values for the FLC ae: $\xi_1 = 1 + 0 + 0 + 1 = 0$ (mod 2), $\xi_2 = 0 + 0 + 0 + 0 = 0$ → The FLC "ae" belongs to the PB. Similarly, you may verify that the FLCs bef, cef, and abd also belong to the PB. By now, we have identified sufficient number of FLCs that belong to the PB because the remaining 11 FLCs in the PB can be obtained from these 4 independent FLCs (ae, bef, cef, abd) by multiplying them (mod 2) two at time , three at a time, and all four at a time. That is to say, the 6th FLC in the PB is (ae)(bef) = $abe^2f = abe^0f$ = abf. The remaining 9 FLCs are: (ae)(cef) = acf, (ae)(abd) = bde, (bef)(cef) = bc, (bef)(abd) = adef, (cef)(abd) = abcdef, (ae)(bef)(cef) = $abce^3$ = abce (mod 2), (ae)(bef)(abd) = df, (ae)(cef)(abd) = bcdf, (bef)(cef)(abd) = acd, and (ae)(bef)(cef)(abd) = cde. The PB is summarized below.

The block with $\xi_1 = \xi_2 = 0$: (1), ae, bef, cef, abd, abf, acf, bde, bc,
The PB: adef, abcdef, abce, df, bcdf, acd, cde

You should check the fact that the values of the contrast functions $\xi_1 = x_1 + x_2 + x_3 + x_5$ and $\xi_2 = x_2 + x_3 + x_4 + x_6$ are indeed zero for every FLC in the above PB.

<u>Step 5.</u> Use the PB to generate the other 3 blocks. Suppose we wish to generate the 16 FLCs in the ($\xi_1 = 0$, $\xi_2 = 1$) block; all we need is to identify only 1 FLC that belongs to this block, say the FLC d = (0, 0, 0, 1, 0, 0). Then we multiply the FLC "d" by the last 15 FLCs of the PB (Mod 2) to generate the ($\xi_1 = 0$, $\xi_2 = 1$) block, which is listed below:

The ($\xi_1 = 0$, $\xi_2 = 1$) block: d, ade, bdef, cdef, ab, abdf, acdf, be, bcd, aef,
abcef, abcde, f, bcf , ac, ce

<u>Exercise 103.</u> Use the above procedure to generate the other 2 blocks of the above 2^{6-2}_{IV} FFD, i.e., generate the ($\xi_1 = 1$, $\xi_2 = 0$) and ($\xi_1 = 1$, $\xi_2 = 1$) blocks.

Fortunately, there is a much simpler method of obtaining the PB of any FFD (in base-2, base-3, base-5, etc), which I describe below in stepwise fashion.

<u>Step I.</u> For the 2^{k-p} FFD, k–p columns can always be written arbitrarily. For the FFD in the injection molding operation, k–p = 6 – 2 = 4 which implies we can write 4 arbitrary columns as shown in an Excel format on my website. However, the assignment of the factors to the 4 columns must be done with extreme care. The first 3 column assignments are again completely arbitrary, but the 4th column (i.e., D) must be done by examining all the generators. Recall that g_1 = ABCE so that the levels of E must be obtained from the product E = +A×B×C, and the fact that g_2 = BCDF, then the levels of factor F must be obtained from the levels of factors B, C, and D as in F = +B×C×D. This leaves only factor D that can be assigned to column 4. If these assignments do produce the PB, then we are almost finished because the other $2^p - 1$ blocks can be generated just like step 5 above.

<u>Step II.</u> If step I does not produce the PB then we have to redo the column assignments as described in the following example.

Example 44. **Consider a 2^{5-2} FFD with the 2 independent generators g_1 = ABC and g_2 = BDE. Since there are 4 blocks containing 8 FLCs each, then we need to sacrifice 3 df and thus we need one more generator namely g_3 = (ABC)(BDE) = ACDE. Further, k – p = 3 implies that we can write 3 columns arbitrarily, bearing in mind that C = A×B because one identity element is I = ABC. The column assignments are shown in Table 12 below.**

Table 12

A	B	D	C=+A×B	E=+ B×D	FLC	y_{ijklm} (LTB)
-	-	-	+	+	ce	14
-	-	+	+	-	cd	35
-	+	-	-	-	b	32
-	+	+	-	+	bde	12
+	-	-	-	+	ae	9
+	-	+	-	-	ad	18
+	+	-	+	-	abc	5
+	+	+	+	+	abcde	7 / 132

Clearly, the above (1/4)th fraction is not the PB because the FLC (1) = (0, 0, 0, 0, 0) = (-1, -1, -1, -1, -1) is not in this block. The above block is actually the ($\xi_1 = x_1 + x_2 + x_3$ = 1, $\xi_2 = x_2 + x_4 + x_5$ = 1) block.

Therefore, we have to redo the C and E columns above in such a manner that both their signs in the 1st run are –1 instead of +1. This implies that we must let C = – A×B in column 4, and E = – B×D in column 5 in order to obtain the PB. Table 13 shows how to generate the PB for this 2^{5-2}_{III} FFD.

Exercise 104. **Use the above procedure to obtain the other 2 blocks of the above 2^{5-2}_{III} FFD and give the corresponding values of their cfs. (b) Determine the alias structure of this FFD. Do aliases change from one block to the next?**

Exercise 104. **Use the above procedure to obtain the other 2 blocks of the above 2^{5-2}_{III} FFD and give the corresponding values of their cfs. (b) Determine the alias structure of this FFD. Do aliases change from one block to the next?**

We now analyze the data of the Example 44 in Table 12 on the previous page of this course notes. It is always best to make a response table (RT) to analyze data from a FFD. This is simply a summary of subtotals (or contrasts) at different levels of a factor (or an effect). The experimenter, however, must be careful to account for the exact number of df that the design matrix provides.

Table 13 (1/4th fraction of a 2^5 factorial with g_1=ABC and g_2=BDE)

A	B	D	C = – A×B	E = – B×D	FLC
-	-	-	-	-	(1)
-	-	+	-	+	de
-	+	-	+	+	bce
-	+	+	+	-	bcd
+	-	-	+	-	ac
+	-	+	+	+	acde
+	+	-	-	+	abe
+	+	+	-	-	abd

This implies that we must list seven 1-df effects in the RT because Table 12 has 8 FLCs and hence provides 7 df for studying effects. The RT of the data of Table 12 is provided below in Table 14. This Table clearly shows that the strongest effect is factor A, and the 2nd most influential effect on the response Y is E, and effects C, D, AD, and CE are relatively weak so that these weak effects should be left out of considerations when obtaining the best FLC in order to maximize Y because Y is an LTB (Larger-The-Better) type response.

Since the AE = CD interactions are strong, they must both be analyzed to determine the optimum levels of (A, E) and (C, D) simultaneously. Therefore, we cross factor A with E, and factor C with D as shown below.

The A×E interaction table atop the next page clearly indicates that the cell (–1, –1) is optimum for A and E, while the C×D Table shows that either the cell (–1, –1) or (+1, +1) is optimum for C and D. Further, the RT 14 shows that the low level of B , –1, is optimum. Hence, the optimal FLC is $X_O = A_0B_0C_1D_1E_0$ = cd, or X_O = (1).

<u>Exercise 105.</u> (a) For the Table 12 on page 180, compute SS(A) and

SS(AE) = SS(CD) in 2 different ways. (b) Obtain the ANOVA Table and verify that all your SS's are correct. ANS: SS(A) = 364.50, SS_C = 12.50.

<u>Exercise 106.</u> Analyze the data of Example 11.17 on pages 438-439 of your text in order to estimate the optimal conditions X_O. That is, obtain the RT and the ANOVA Table first, and then identify X_O.

<u>Exercise 107.</u> Work Exercises 49 on page 442 of Devore's 7th edition and obtain X_O.

Table 14 (The RT for the data of Table 12)

Effects	**A**	**B**	**C**	**D**	**E**	**AD = CE**	**AE = CD**
ℓ_0	**93**	**76**	**71**	**60**	**90**	**61**	**49**
ℓ_1	**39**	**56**	**61**	**72**	**42**	**71**	**83**
$\ell_1 - \ell_0$	**- 54**	**- 20**	**- 10**	**12**	**- 48**	**10**	**34**
Rank of $\lvert \ell_1 - \ell_0 \rvert$	**1**	**4**	**6.5**	**5**	**2**	**6.5**	**3**

The A×E Table

E A	-1	+1
- 1	67*	26
+1	23	16

The C×D Table

D C	-1	+1
- 1	41*	30
+1	19	42*

TEST 3