4 Two-Level (2^k) Factorial Designs

- Many applications of response surface methodology are based on fitting one of the following models:

$$\text{First order model} \quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \tag{3}$$

$$\text{Interaction model} \quad y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum\sum_{i<j}^{k} \beta_{ij} x_i x_j \tag{4}$$

$$\text{Second order model} \quad y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum\sum_{i<j}^{k} \beta_{ij} x_i x_j + \sum_{i=1}^{k} \beta_{ii} x_i^2 \tag{5}$$

- One commonly-used response surface design is a 2^k *factorial design.*
- A $\mathbf{2^k}$ **factorial design** is a k-factor design such that
  (i) Each factor has two levels (coded -1 and $+1$).
  (ii) The 2^k experimental runs are based on the 2^k combinations of the ± 1 factor levels.
- Common applications of 2^k factorial designs (and the fractional factorial designs in Section 5 of the course notes) include the following:
  - *As screening experiments*: A 2^k design is used to identify or **screen** for potentially important process or system variables. Once screened, these important variables are then incorporated into a more complex experimental study.
  - *To fit the first-order model in (3) or the interaction model in (4)*: The 2^k design can be used to fit model (3) or (4). One application of fitting these models is in the method of steepest ascent or descent (Section 6 of the course notes).
  - *As a building block for second-order response surface designs*: 2^k designs are used to generate central composite designs (CCDs) and Box-Behnken designs (BBDs).
- We will first analyze each 2^k design as a *fixed effects* design. We will also generalize the fixed effects results to the regression model approach for which the model contains regression coefficients $\beta_0, \beta_1, \beta_2, \ldots$ as in (3) and (4).
- Before analyzing the data, you must determine if the design was completely randomized or if blocking was used. Your answer to this question will indicate the appropriate analysis. Initially, we will assume the design was completely randomized.

4.1 The 2^2 Design

- The simplest 2^k design is the 2^2 design. This is a special case of a two-factor factorial design with factors A and B having two levels.
- Because a 2^2 design has only 4 runs, several (n) replications are taken.
- Notationally, we use lowercase letters a, b, ab, and (1) to indicate the sum of the responses for all replications at each of the corresponding levels of A and B.
  - If the lower case letter appears, then that factor is at its high (+1) level.
  - If the lower case letter does not appear, then that factor is at its low (-1) level.

Factor Level Combination	Coded Levels	Replicate 1	2	$\cdots$	n	Sum of n Replicates	
A low , B low	-1 -1	xxx	xxx	$\cdots$	xxx	(1)	$= y_{11\cdot}$
A high, B low	$+1$ -1	xxx	xxx	$\cdots$	xxx	a	$= y_{21\cdot}$
A low , B high	-1 $+1$	xxx	xxx	$\cdots$	xxx	b	$= y_{12\cdot}$
A high, B high	$+1$ $+1$	xxx	xxx	$\cdots$	xxx	ab	$= y_{22\cdot}$

- We will use the notation A^+ and A^- to represent the set of observations with factor A at its high $(+1)$ and its low (-1) levels, respectively. The same notation applies to B^+ and B^- for factor B.

  a and ab correspond to A^+ and (1) and b correspond to A^-.

  b and ab correspond to B^+ and (1) and a correspond to B^-.

- $\overline{y}_{A^+}$ and $\overline{y}_{A^-}$ are the means of all observations when $A = +1$ and $A = -1$, respectively.
- $\overline{y}_{B^+}$ and $\overline{y}_{B^-}$ are the means of all observations when $B = +1$ and $B = -1$, respectively.
- The **average effect of a factor** is the average change in the response produced by a change in the level of that factor *averaged over the levels of the other factor.*
- For a 2^2 design with n replicates, the

— **Average effect of Factor *A***, denoted A, is

$$A \;=\; \overline{y}_{A^+} - \overline{y}_{A^-} \;= \qquad\qquad = \frac{1}{2n}\left[ab + a - b - (1)\right].$$

— **Average effect of Factor *B***, denoted B, is

$$B \;=\; \overline{y}_{B^+} - \overline{y}_{B^-} \;= \qquad\qquad = \frac{1}{2n}\left[ab - a + b - (1)\right].$$

— **Interaction effect between Factors *A* and *B***, denoted AB, is the difference between (i) the average change in response when the levels of Factor A are changed given Factor B is at its high level and (ii) the average change in response when the levels of Factor A are changed given Factor B is at its low level:

$$\begin{aligned} AB &= (\overline{y}_{A^+B^+} - \overline{y}_{A^-B^+}) \;-\; (\overline{y}_{A^+B^-} - \overline{y}_{A^-B^-}) \\ &= \qquad\qquad = \frac{ab - a - b + (1)}{2n} \end{aligned}$$

Note: The results would be the same if we switched the roles of A and B in the definition:

$$\begin{aligned} AB &= (\overline{y}_{A^+B^+} - \overline{y}_{A^+B^-}) \;-\; (\overline{y}_{A^-B^+} - \overline{y}_{A^-B^-}) \\ &= \qquad\qquad = \frac{ab - a - b + (1)}{2n} \end{aligned}$$

Sums of Squares for A, B and AB.

- Note that when estimating the effects for A, B and AB the following contrasts are used:

$$\Gamma_A = ab + a - b - (1) \qquad \Gamma_B = ab - a + b - (1) \qquad \Gamma_{AB} = ab - a - b + (1)$$

- Γ_A, Γ_B, and Γ_{AB} are used to estimate A, B, and AB, and they are **orthogonal contrasts**.
  - The coefficient vectors for the contrasts are $[1 \;\; 1 -1 -1]$ for A, $[1 -1 \;\; 1 -1]$ for B, and $[1 -1 -1 \;\; 1]$ for AB. Note the dot product of any two vectors = 0. This is why they are called orthogonal contrasts.

- The sum of squares for contrast Γ is 7

- For a replicated 2^2 design, this is equivalent to:

$$SS_A = \frac{[ab + a - b - (1)]^2}{4n} \qquad SS_B = \frac{[ab - a + b - (1)]^2}{4n} \qquad SS_{AB} = \frac{[ab - a - b + (1)]^2}{4n}$$

- Because there are two levels for both factors, the degrees of freedom associated with each sum of squares is 1. Thus, $MS_A = SS_A$, $MS_B = SS_B$, and $MS_{AB} = SS_{AB}$.

- Because there are n replicates for each of the four $A * B$ treatment combinations, there are $4(n-1)$ degrees of freedom for error for the four-parameter interaction model in (4).

- It is common to list the treatment combinations in **standard order**: (1), a, b, and ab. Many references use a shortened notation ($-$ or $+$) to denote the low (-1) and high ($+1$) levels of a factor.

Example: An engineer designs a 2^2 design with $n = 4$ replicates to study the effects of bit size (A) and cutting speed (B) on routing notches in a printed circuit board.

A	B	AB	Replicates				Totals
$-$	$-$	$+$	18.2	18.9	12.9	14.4	$(1) = 64.4$
$+$	$-$	$-$	27.2	24.0	22.4	22.5	$a = 96.1$
$-$	$+$	$-$	15.9	14.5	15.1	14.2	$b = 59.7$
$+$	$+$	$+$	41.0	43.9	36.3	39.9	$ab = 161.1$

Note: the signs in the AB column are the signs that result when multiplying the A and B columns.

- The estimates of the fixed effects are:

$$\begin{aligned} A &= \frac{\Gamma_A}{2n} = \frac{ab + a - b - (1)}{2n} = \frac{161.1 + 96.1 - 59.7 - 64.4}{8} = \\ B &= \frac{\Gamma_B}{2n} = \frac{ab - a + b - (1)}{2n} = \frac{161.1 - 96.1 + 59.7 - 64.4}{8} = \\ AB &= \frac{\Gamma_{AB}}{2n} = \frac{ab - a - b + (1)}{2n} = \frac{161.1 - 96.1 - 59.7 + 64.4}{8} = \end{aligned}$$

- The sum of squares $SS_i = \Gamma_i^2/4n$ for $i = A, B, AB, T$ is:

$$SS_A = \frac{133.1^2}{16} = 1107.2256 \qquad SS_B = \frac{60.3^2}{16} = 227.2556$$

$$SS_{AB} = \frac{69.7^2}{16} = 303.6306 \qquad SS_T = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{4} y_{ijk}^n - \frac{y_{\cdots}^2}{4n} = 10796.7 - \frac{381.3^2}{16} = 1709.8344$$

$$SS_E = SS_T - SS_A - SS_B - SS_{AB} = 71.7225$$

- Sums of squares can also be calculated using the formulas for a two-factor factorial design.

The Regression Model

- If both factors in the 2^2 design are quantitative (say, x_1 and x_2), we can fit the first order regression model
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon.$$
or, we can fit the regression model with interaction:
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \epsilon.$$

- The least squares estimates $[\, b_0\ b_1\ b_2\ b_{12}\,]' = (\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ are directly related to the estimated effects A, B, and AB from the fixed effects analysis:
$$\begin{aligned} b_0 &= \frac{ab + a + b + (1)}{4n} && \text{or} \quad b_0 = \overline{y} \\ b_1 &= \frac{\Gamma_A}{4n} = \frac{ab + a - b - (1)}{4n} && \text{or} \quad b_1 = A/2 \\ b_2 &= \frac{\Gamma_B}{4n} = \frac{ab + b - a - (1)}{4n} && \text{or} \quad b_2 = B/2 \\ b_{12} &= \frac{\Gamma_{AB}}{4n} = \frac{ab + (1) - a - b}{4n} && \text{or} \quad b_2 = AB/2 \end{aligned}$$

- For the previous example:
$$\begin{aligned} b_0 &= \overline{y} = 381.3/16 = 23.83125 \\ b_1 &= A/2 = 16.6375/2 = 8.31875 \\ b_2 &= B/2 = 7.5375/2 = 3.76875 \\ b_{12} &= AB/2 = 8.7125/2 = 4.35625 \end{aligned}$$

- Therefore, the fitted regression equation is
$$\widehat{y} = 23.83125 + 8.31875 x_1 + 3.76875 x_2 + 4.35625 x_1 x_2$$
where (x_1, x_2) are the coded levels of factors A and B.

4.2 The 2^3 Design

- Let A, B, and C be three factors each having two levels. The design which includes the $2^3 = 8$ treatment combinations of $A * B * C$ is called a 2^3 **(factorial) design**.

- The following table summarizes the eight treatment combinations and the signs for calculating effects in the 2^3 design (I =intercept). Assume each treatment is replicted n times.

Factorial Effect								Sum of	
I	A	B	C	AB	AC	BC	ABC	replicates	
+	−	−	−	+	+	+	−	(1)	$= y_{111\cdot}$
+	+	−	−	−	−	+	+	a	$= y_{211\cdot}$
+	−	+	−	−	+	−	+	b	$= y_{121\cdot}$
+	+	+	−	+	−	−	−	ab	$= y_{221\cdot}$
+	−	−	+	+	−	−	+	c	$= y_{112\cdot}$
+	+	−	+	−	+	−	−	ac	$= y_{212\cdot}$
+	−	+	+	−	−	+	−	bc	$= y_{122\cdot}$
+	+	+	+	+	+	+	+	abc	$= y_{222\cdot}$

- The signs in the interaction columns are the signs that result when multiplying the main effect columns in the interaction of interest. Note that all columns are mutually orthogonal.

- For a 2^3 design with n replicates, each estimated effect is the differences between two means: The first mean is the average of all data corresponding to the + rows in an effect column and the second mean is the average of all data corresponding to the $-$ rows in an effect column.

  Average effect of Factor *A*, denoted A, is

$$\begin{aligned} A &= \overline{y}_{A^+} - \overline{y}_{A^-} = \frac{(a+ab+ac+abc)}{4n} - \frac{(1)+b+c+bc}{4n} \\ &= \frac{1}{4n}\left[a+ab+ac+abc-(1)-b-c-bc\right]. \end{aligned}$$

  Average effect of Factor *B*, denoted B, is

$$\begin{aligned} B &= \overline{y}_{B^+} - \overline{y}_{B^-} = \frac{(b+ab+bc+abc)}{4n} - \frac{(1)+a+c+ac}{4n} \\ &= \frac{1}{4n}\left[b+ab+bc+abc-(1)-a-c-ac\right]. \end{aligned}$$

  Average effect of Factor *C*, denoted C, is

$$\begin{aligned} C &= \overline{y}_{C^+} - \overline{y}_{C^-} = \frac{(c+ac+bc+abc)}{4n} - \frac{(1)+a+b+ab}{4n} \\ &= \frac{1}{4n}\left[c+ac+bc+abc-(1)-a-b-ab\right]. \end{aligned}$$

  Two-factor interaction effect between Factors *A* and *B*, denoted AB, is

$$AB = \frac{ab+abc-a-ac}{4n} - \frac{b+bc-(1)-c}{4n} = \frac{abc+ab+c+(1)-a-ac-bc-b}{4n}.$$

  Two-factor interaction effect between Factors *A* and *C*, denoted AC, is

$$AC = \frac{ac+abc-a-ab}{4n} - \frac{c+bc-(1)-b}{4n} = \frac{abc+ac+b+(1)-ab-a-bc-c}{4n}.$$

  Two-factor interaction effect between Factors *B* and *C*, denoted BC, is

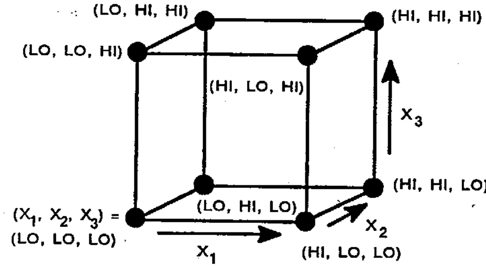
$$BC = \frac{bc+abc-b-ab}{4n} - \frac{c+ac-(1)-a}{4n} = \frac{abc+bc+a+(1)-ab-b-ac-c}{4n}.$$

  Three-factor interaction effect between Factors *A*, *B* and *C*, denoted ABC, is the average difference between the AB interaction for the two different levels of C. That is,

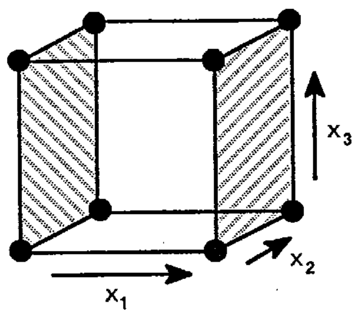
$$\begin{aligned} ABC &= \frac{(abc-bc)-(ac-c)}{4n} - \frac{(ab-b)-(a-(1))}{4n} \\ &= \frac{abc+a+b+c-ab-ac-bc-(1)}{4n} \end{aligned}$$

- Let Γ = the contrast sum in the numerator for any of the effects. Then the sums of squares associated with that effect is $SS =$

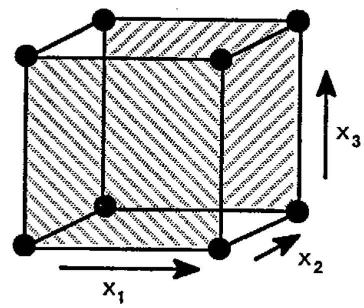
Geometric Representation for a 2^3 Design



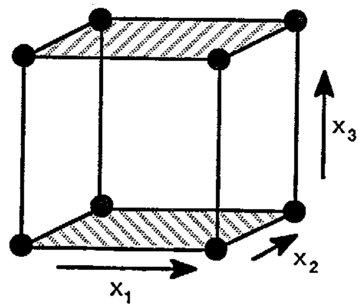
A effect



B effect

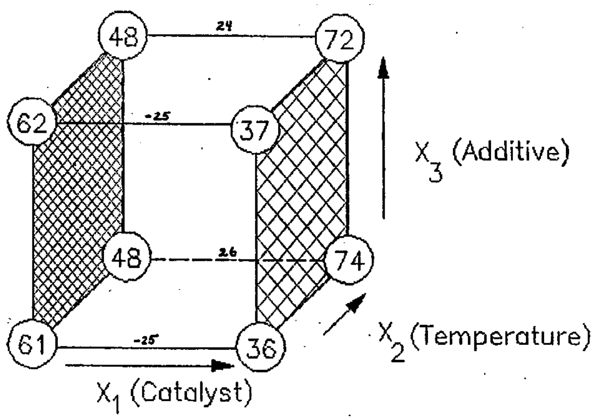


C effect



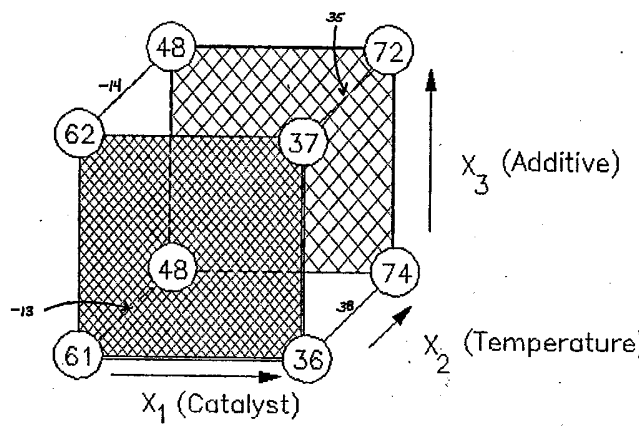
Estimation of Main Effects

A effect



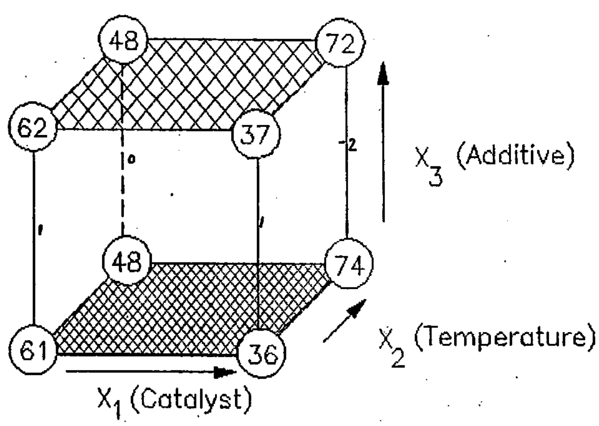
$$\frac{72+37+74+36}{4} - \frac{48+62+48+61}{4}$$

B effect



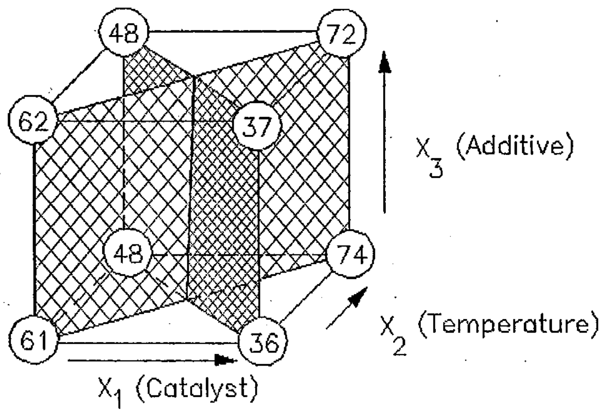
$$\frac{48+72+48+74}{4} - \frac{62+37+61+36}{4}$$

C effect

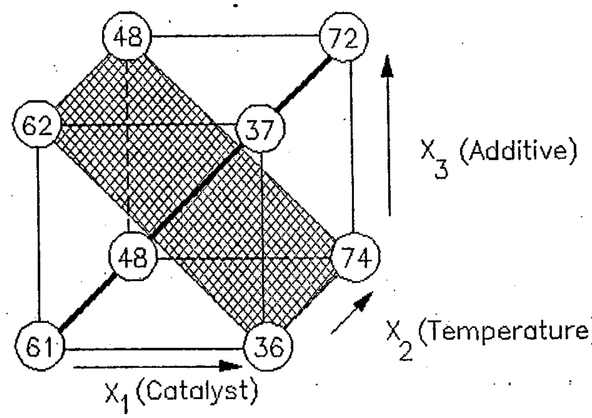


$$\frac{48+72+62+37}{4} - \frac{48+74+61+36}{4}$$

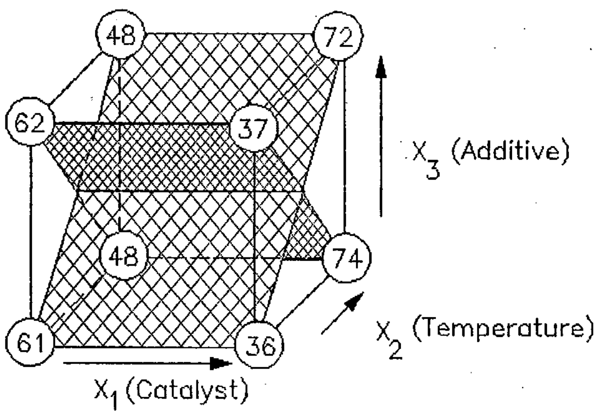
Estimation of Two-Factor Interaction Effects



$$\frac{72+74+62+61}{4} - \frac{48+37+48+36}{4} = 25$$

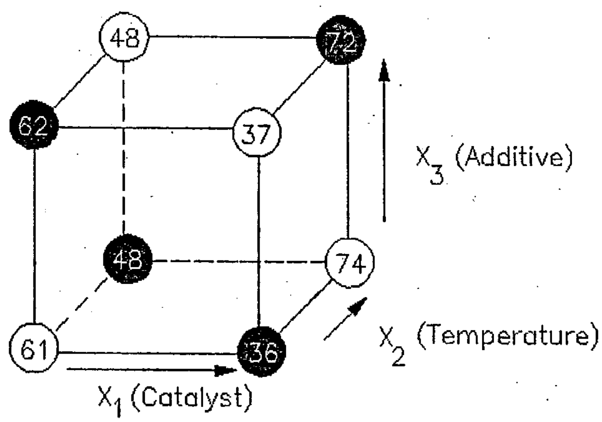


$$\frac{37+72+48+61}{4} - \frac{48+62+74+36}{4} = -0.5$$



$$\frac{48+72+61+36}{4} - \frac{62+37+48+74}{4} = -1$$

Estimation of the Three-Factor Interaction Effect



$$\frac{72+62+48+36}{4} - \frac{48+37+74+61}{4} = -0.5$$

The Regression Model

- If all three factors in the 2^3 design are quantitative (say, x_1, x_2, and x_3), we can fit the regression model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3 + \epsilon. \quad (6)$$

- The least squares estimates (with the exception of b_0) are 1/2 of the estimated effects from the fixed effects analysis. That is,

$$b_0 = \overline{y} \qquad b_1 = A/2 \qquad b_2 = B/2 \qquad b_3 = C/2$$

$$b_{12} = AB/2 \qquad b_{13} = AC/2 \qquad b_{23} = BC/2 \qquad b_{123} = ABC/2$$

- Because all of the contrasts associated with each of the effects are orthogonal, the least squares estimates remain unchanged for any model containing a subset of terms in (6).

4.2.1 A 2^3 Design Example

An engineer is interested in the effects of cutting speed (A), tool geometry (B), and cutting angle (C) on the life (in hours) of a machine tool. Two levels of each factor are chosen, and three replicates of a 2^3 design are run. The results are summarized below:

A x_1	B x_2	C x_3	Replicates			Treatment Sums
$-$	$-$	$-$	22	31	25	$(1) = 78$
$+$	$-$	$-$	32	43	29	$a = 104$
$-$	$+$	$-$	35	34	50	$b = 119$
$+$	$+$	$-$	55	47	46	$ab = 148$
$-$	$-$	$+$	44	45	38	$c = 127$
$+$	$-$	$+$	40	37	36	$ac = 113$
$-$	$+$	$+$	60	50	54	$bc = 164$
$+$	$+$	$+$	39	41	47	$abc = 127$

Analyze the data (with lack-of-fit tests) assuming the following 4 models:

- (Model 1): An additive model with fixed (categorical) effects.
- (Model 2): A first-order regression model.
- (Model 3): An interaction model with fixed (categorical) effects.
- (Model 4): A regression model with all two-factor crossproduct (interaction). terms.

Note there are _______ df for pure error.

- We will first estimate effects and sums of squares using the formulas, then use SAS to perform the analysis. Recall:

(1)	a	b	ab	c	ac	bc	abc
78	104	119	148	127	113	164	127

Model									
Fixed Effects $\longrightarrow$	I	A	B	C	AB	AC	BC	ABC	Treatment
Regression $\longrightarrow$	Int	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	Sums
	+	−	−	−	+	+	+	−	$(1) = 78$
	+	+	−	−	−	−	+	+	$a = 104$
	+	−	+	−	−	+	−	+	$b = 119$
	+	+	+	−	+	−	−	−	$ab = 148$
	+	−	−	+	+	−	−	+	$c = 127$
	+	+	−	+	−	+	−	−	$ac = 113$
	+	−	+	+	−	−	+	−	$bc = 164$
	+	+	+	+	+	+	+	+	$abc = 127$

- The fixed effects estimates are

$$\begin{aligned}
A &= \frac{104+148+113+127-78-119-127-164}{(4)(3)} = \frac{4}{12} = .\overline{3} \\
B &= \frac{119+148+164+127-78-104-127-113}{(4)(3)} = \frac{136}{12} = 11.\overline{3} \\
C &= \frac{127+113+164+127-78-104-119-148}{(4)(3)} = \frac{82}{12} = 6.8\overline{3} \\
AB &= \frac{78+148+127+127-104-119-113-164}{(4)(3)} = \frac{-20}{12} = -1.\overline{6} \\
AC &= \frac{78+119+113+127-104-148-127-164}{(4)(3)} = \frac{-106}{12} = -8.8\overline{3} \\
BC &= \frac{78+104+164+127-119-148-127-113}{(4)(3)} = \frac{-34}{12} = -2.8\overline{3} \\
ABC &= \frac{104+119+127+127-78-148-113-164}{(4)(3)} = \frac{-26}{12} = -2.1\overline{6}
\end{aligned}$$

- The sums of squares are calculated using $\dfrac{\Gamma^2_{effect}}{8n}$:

$$SS_A = \frac{4^2}{24} = .\overline{6} \qquad SS_B = \frac{(136)^2}{24} = 770.\overline{6} \qquad SS_C = \frac{82^2}{24} = 280.1\overline{6}$$

$$SS_{AB} = \frac{(-20)^2}{24} = 16.\overline{6} \qquad SS_{AC} = \frac{(-106)^2}{24} = 468.1\overline{6}$$

$$SS_{BC} = \frac{(-34)^2}{24} = 48.1\overline{6} \qquad SS_{ABC} = \frac{(-26)^2}{24} = 28.1\overline{6}$$

- Fixed effects additive model (Model 1):

$$y_{ijkl} \;=\; \mu \;+\; \alpha_i \;+\; \beta_j \;+\; \gamma_k \;+\; \epsilon_{ijkl} \qquad (i = \pm 1,\ j = \pm 1,\ k = \pm 1,\ l = 1, 2, 3)$$

- Note the effect estimates in the SAS output match the formula calculations.
- First-order regression model (Model 2): For $i = 1, 2, \ldots, 24$

$$y_i \;=\; \beta_0 \;+\; \beta_1 x_{1i} \;+\; \beta_2 x_{2i} \;+\; \beta_3 x_{3i} + \epsilon_i$$

  Note that the parameter estimates are 1/2 of those from the fixed effects in Model 1.

- For Models 1 and 2, there are df for pure error and df for total error. Thus, the df for lack-of-fit = . This means we can add at most additional terms in the model (such as interaction terms).
- There is a significant lack-of-fit (p-value =). We can add at most additional terms in the model (such as interaction terms).
- The residuals in the Residual vs Predicted Value plot (page 50) are not randomly scattered about 0 for several (x_1, x_2, x_3) combinations. This suggests a lack-of-fit problem.

MODEL 1: ADDITIVE FIXED EFFECTS MODEL

The GLM Procedure

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1051.500000	350.500000	6.72	0.0026
Error	20	1043.833333	52.191667		
Corrected Total	23	2095.333333			

R-Square	Coeff Var	Root MSE	Y Mean
0.501829	17.69236	7.224380	40.83333

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	1	0.6666667	0.6666667	0.01	0.9111
B	1	770.6666667	770.6666667	14.77	0.0010
C	1	280.1666667	280.1666667	5.37	0.0312

Parameter	Estimate	Standard Error	t Value	Pr > |t|
A	0.3333333	2.94934079	0.11	0.9111
B	11.3333333	2.94934079	3.84	0.0010
C	6.8333333	2.94934079	2.32	0.0312

MODEL 2: FIRST ORDER REGRESSION MODEL

The REG Procedure
Model: MODEL1
Dependent Variable: Y

Number of Observations Read	24
Number of Observations Used	24

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1051.50000	350.50000	6.72	0.0026
Error	20	1043.83333	52.19167		
Lack of Fit	4	561.16667	140.29167	4.65	0.0111
Pure Error	16	482.66667	30.16667		
Corrected Total	23	2095.33333			

Root MSE	7.22438	R-Square	0.5018
Dependent Mean	40.83333	Adj R-Sq	0.4271
Coeff Var	17.69236		

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	40.83333	1.47467	27.69	<.0001	0
X1	1	0.16667	1.47467	0.11	0.9111	1.00000
X2	1	5.66667	1.47467	3.84	0.0010	1.00000
X3	1	3.41667	1.47467	2.32	0.0312	1.00000

Level of A	N	Y Mean	Y Std Dev
-1	12	40.6666667	11.7808267
1	12	41.0000000	7.1858447

Level of B	N	Y Mean	Y Std Dev
-1	12	35.1666667	7.46912838
1	12	46.5000000	8.03967435

Level of C	N	Y Mean	Y Std Dev
-1	12	37.4166667	10.5093753
1	12	44.2500000	7.3870279

Level of A	Level of B	N	Y Mean	Y Std Dev
-1	-1	6	34.1666667	9.7039511
-1	1	6	47.1666667	10.4769588
1	-1	6	36.1666667	5.1153364
1	1	6	45.8333333	5.6005952

Level of A	Level of C	N	Y Mean	Y Std Dev
-1	-1	6	32.8333333	9.82683401
-1	1	6	48.5000000	7.84219357
1	-1	6	42.0000000	9.79795897
1	1	6	40.0000000	3.89871774

Level of B	Level of C	N	Y Mean	Y Std Dev
-1	-1	6	30.3333333	7.25718035
-1	1	6	40.0000000	3.74165739
1	-1	6	44.5000000	8.36062199
1	1	6	48.5000000	7.91833316

- Now let's add the three two-factor interactions to get Models 3 and 4.
- Fixed effects interaction model (Model 3):

$$y_{ijkl} \;=\; \mu \;+\; \alpha_i \;+\; \beta_j \;+\; \gamma_k \;+\; \alpha\beta_{ij} \;+\; \alpha\gamma_{ik} \;+\; \beta\gamma_{jk} \;+\; \epsilon_{ijkl}$$

for $(i = \pm 1,\ j = \pm 1,\ k = \pm 1,\ l = 1, 2, 3)$

- Note the effect estimates match the formula calculations.
- Interaction regression model (Model 4): For $i = 1, 2, \ldots, 24$

$$y_i \;=\; \beta_0 \;+\; \beta_1 x_{1i} \;+\; \beta_2 x_{2i} \;+\; \beta_3 x_{3i} + \;+\; \beta_{12} x_{1i} x_{2i} \;+\; \beta_{13} x_{1i} x_{3i} \;+\; \beta_{23} x_{2i} x_{3i} \;+\; \epsilon_i$$

Note that the parameter estimates are 1/2 of those from the fixed effects in Model 3.

- The residuals are randomly scattered about 0. This suggests there is no lack-of-fit problem. The lack-of-fit test (p-value=) supports this.

MODEL 3: INTERACTION FIXED EFFECTS MODEL

The GLM Procedure

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	1584.500000	264.083333	8.79	0.0002
Error	17	510.833333	30.049020		
Corrected Total	23	2095.333333			

R-Square	Coeff Var	Root MSE	Y Mean
0.756204	13.42457	5.481699	40.83333

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	1	0.6666667	0.6666667	0.02	0.8833
B	1	770.6666667	770.6666667	25.65	<.0001
A*B	1	16.6666667	16.6666667	0.55	0.4666
C	1	280.1666667	280.1666667	9.32	0.0072
A*C	1	468.1666667	468.1666667	15.58	0.0010
B*C	1	48.1666667	48.1666667	1.60	0.2226

Parameter	Estimate	Standard Error	t Value	Pr > |t|
A	0.3333333	2.23789408	0.15	0.8833
B	11.3333333	2.23789408	5.06	<.0001
C	6.8333333	2.23789408	3.05	0.0072
A*B	-1.6666667	2.23789408	-0.74	0.4666
A*C	-8.8333333	2.23789408	-3.95	0.0010
B*C	-2.8333333	2.23789408	-1.27	0.2226

MODEL 4: INTERACTION REGRESSION MODEL

The REG Procedure
Model: MODEL1
Dependent Variable: Y

Number of Observations Read	24
Number of Observations Used	24

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	1584.50000	264.08333	8.79	0.0002
Error	17	510.83333	30.04902		
Lack of Fit	1	28.16667	28.16667	0.93	0.3483
Pure Error	16	482.66667	30.16667		
Corrected Total	23	2095.33333			

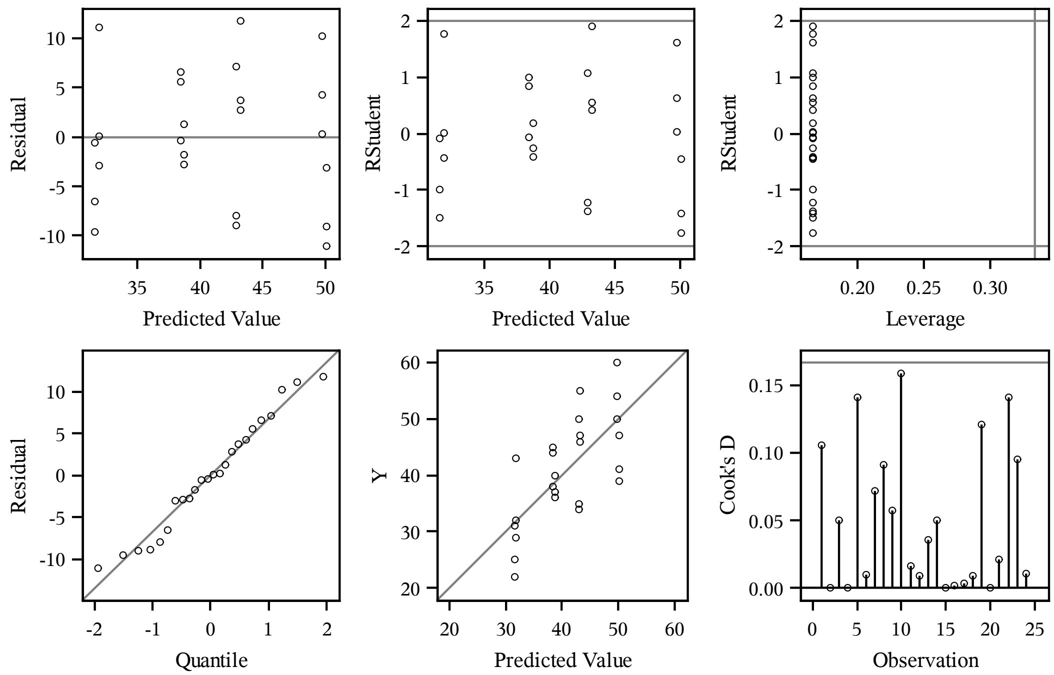
Root MSE	5.48170	R-Square	0.7562
Dependent Mean	40.83333	Adj R-Sq	0.6702
Coeff Var	13.42457		

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	40.83333	1.11895	36.49	<.0001	0
X1	1	0.16667	1.11895	0.15	0.8833	1.00000
X2	1	5.66667	1.11895	5.06	<.0001	1.00000
X3	1	3.41667	1.11895	3.05	0.0072	1.00000
X1X2	1	-0.83333	1.11895	-0.74	0.4666	1.00000
X1X3	1	-4.41667	1.11895	-3.95	0.0010	1.00000
X2X3	1	-1.41667	1.11895	-1.27	0.2226	1.00000

MODEL 2: FIRST ORDER REGRESSION MODEL

The REG Procedure
Model: MODEL1
Dependent Variable: Y

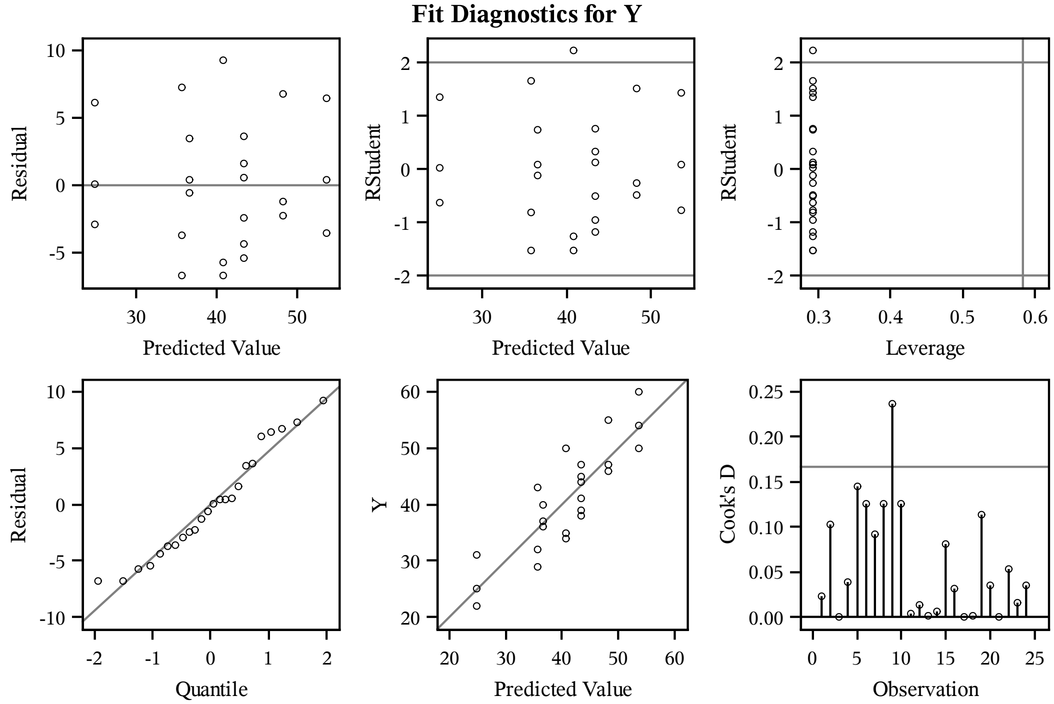
Fit Diagnostics for Y



MODEL 4: INTERACTION REGRESSION MODEL

The REG Procedure
Model: MODEL1
Dependent Variable: Y

Fit Diagnostics for Y



SAS Code for the 2^3 Design Example

- **ESTIMATE** statements in SAS are used to calculate average effect estimates.
- Because of orthogonality, all standard errors are identically

$$2.24227067 \; = \; \sqrt{MSE/2n} \; = \; \sqrt{30.1667/6}$$

```
DM 'LOG; CLEAR; OUT; CLEAR;';

ODS LISTING;
ODS PRINTER PDF file='C:\COURSES\ST578\SAS\TWO3.PDF';
OPTIONS NODATE NONUMBER;

OPTIONS PS=54 LS=76 NODATE NONUMBER;

DATA IN;
   DO C = -1 TO 1 BY 2;
   DO B = -1 TO 1 BY 2;
   DO A = -1 TO 1 BY 2;
   DO REP = 1 TO 3;
      INPUT Y @@;
      X1=A;  X2=B;  X3=C;
      X1X2 = X1*X2;  X1X3 = X1*X3;  X2X3 = X2*X3;
      OUTPUT;
   END; END; END; END;
LINES;
22 31 25   32 43 29   35 34 50   55 47 46
44 45 38   40 37 36   60 50 54   39 41 47
;
PROC GLM DATA=IN PLOTS=NONE;
  CLASS A B C;
  MODEL Y = A B C / SS3;
  MEANS A B C;
  ESTIMATE 'A' A -1 1;
  ESTIMATE 'B' B -1 1;
  ESTIMATE 'C' C -1 1;
TITLE 'MODEL 1: ADDITIVE FIXED EFFECTS MODEL';

PROC REG DATA=IN PLOTS=(DIAGNOSTICS);
     MODEL Y = X1 X2 X3 / LACKFIT VIF;
TITLE 'MODEL 2: FIRST ORDER REGRESSION MODEL';

PROC GLM DATA=IN PLOTS=NONE;
  CLASS A B C;
  MODEL Y = A|B|C@2 / SS3 ;
  MEANS A|B|C@2;
  ESTIMATE 'A' A -1 1;
  ESTIMATE 'B' B -1 1;
  ESTIMATE 'C' C -1 1;
  ESTIMATE 'A*B' A*B 1 -1 -1 1 / DIVISOR=2;
  ESTIMATE 'A*C' A*C 1 -1 -1 1 / DIVISOR=2;
  ESTIMATE 'B*C' B*C 1 -1 -1 1 / DIVISOR=2;
* ESTIMATE 'A*B*C' A*B*C -1 1 1 -1 1 -1 -1 1 ;
TITLE 'MODEL 3: INTERACTION FIXED EFFECTS MODEL';

PROC REG DATA=IN PLOTS=(DIAGNOSTICS);
     MODEL Y = X1 X2 X3 X1X2 X1X3 X2X3 / LACKFIT VIF;
TITLE 'MODEL 4: INTERACTION REGRESSION MODEL';
RUN;
```

4.3 Analyzing Unreplicated Experiments

- To test hypotheses in an unreplicated 2^k design ($n = 1$), it is necessary to "pool" interaction terms (especially higher-order interaction terms), and use the MSE after pooling as an estimate of the random error σ^2.

- The problem is to determine which interaction terms should be pooled together. The following three steps are recommended:

  1. Estimate all effects for the full-factorial interaction model.
  2. Make a normal probability plot of the estimated effects (excluding the intercept), and label the "outlier" effects. Higher-order interactions which are not outliers can be pooled to form the MSE.
  3. Run the ANOVA using this pooled error term.

- Warning: When a higher-order interaction exists, it is inappropriate to pool that interaction with the other interactions because it will inflate the MSE.

- Some comments on the normal probability plot of the $2^k - 1$ estimates for either the fixed effects or regression model:

  – If an effect is not significantly different than zero, then it should be randomly and normally distributed about 0. That is, it is $N(0, \sigma^2/\quad$. When plotted, all of the effects which are not significantly different than zero should lie along a straight line on the normal probability plot.

  – If an effect is significantly different than zero, then it should be randomly and normally distributed about its mean which we will call β. That is, the effect is $N(\beta, \sigma^2/\quad)$. Then, in the normal probability plot, all of the non-zero effects will be plotted away from the line formed by the zero-mean effects.

Unreplicated 2^4 Design Example (from Montgomery text): In a process development study on process yield in pounds, four factors were studied: time, concentration (conc), pressure , and temperature (temp). Each factor had two levels. A single replicate of the 2^4 design was run as a completely randomized design. The resulting data are shown in the following table:

time	conc	pressure	temp	yield
−	−	−	−	12
+	−	−	−	18
−	+	−	−	13
+	+	−	−	16
−	−	+	−	17
+	−	+	−	15
−	+	+	−	20
+	+	+	−	15
−	−	−	+	10
+	−	−	+	25
−	+	−	+	13
+	+	−	+	24
−	−	+	+	19
+	−	+	+	21
−	+	+	+	17
+	+	+	+	23

Analyze the data from this unreplicated experiment from *Design and Analysis of Experiments*, by D. Montgomery (8th ed., p.298).

A 2**4 DESIGN -- ESTIMATION OF EFFECTS

The GLM Procedure

Dependent Variable: YIELD

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	15	291.7500000	19.4500000	.	.
Error	0	0.0000000	.		
Corrected Total	15	291.7500000			

R-Square	Coeff Var	Root MSE	YIELD Mean
1.000000	.	.	17.37500

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TIME	1	81.00	81.00	.	.
CONC	1	1.00	1.00	.	.
TIME*CONC	1	2.25	2.25	.	.
PRESSURE	1	16.00	16.00	.	.
TIME*PRESSURE	1	72.25	72.25	.	.
CONC*PRESSURE	1	0.25	0.25	.	.
TIME*CONC*PRESSURE	1	4.00	4.00	.	.
TEMP	1	42.25	42.25	.	.
TIME*TEMP	1	64.00	64.00	.	.
CONC*TEMP	1	0.00	0.00	.	.
TIME*CONC*TEMP	1	2.25	2.25	.	.
PRESSURE*TEMP	1	0.00	0.00	.	.
TIME*PRESSURE*TEMP	1	0.25	0.25	.	.
CONC*PRESSURE*TEMP	1	2.25	2.25	.	.
TIME*CONC*PRESS*TEMP	1	4.00	4.00	.	.

	Parameter	Estimate	Standard Error	t Value	Pr > |t|
A	TIME	4.50	.	.	.
B	CONC	0.50	.	.	.
C	PRESSURE	2.00	.	.	.
D	TEMP	3.25	.	.	.
A*B	TIME*CONC	-0.75	.	.	.
A*C	TIME*PRES	-4.25	.	.	.
A*D	TIME*TEMP	4.00	.	.	.
B*C	CONC*PRES	0.25	.	.	.
B*D	CONC*TEMP	0.00	.	.	.
C*D	PRES*TEMP	0.00	.	.	.
A*B*C	TIME*C*P	1.00	.	.	.
A*B*D	TIME*C*T	0.75	.	.	.
A*C*D	TIME*P*T	-0.25	.	.	.
B*C*D	C*P*TEMP	-0.75	.	.	.
A*B*C*D	T*C*P*T	1.00	.	.	.

^^^^^^^^^^^

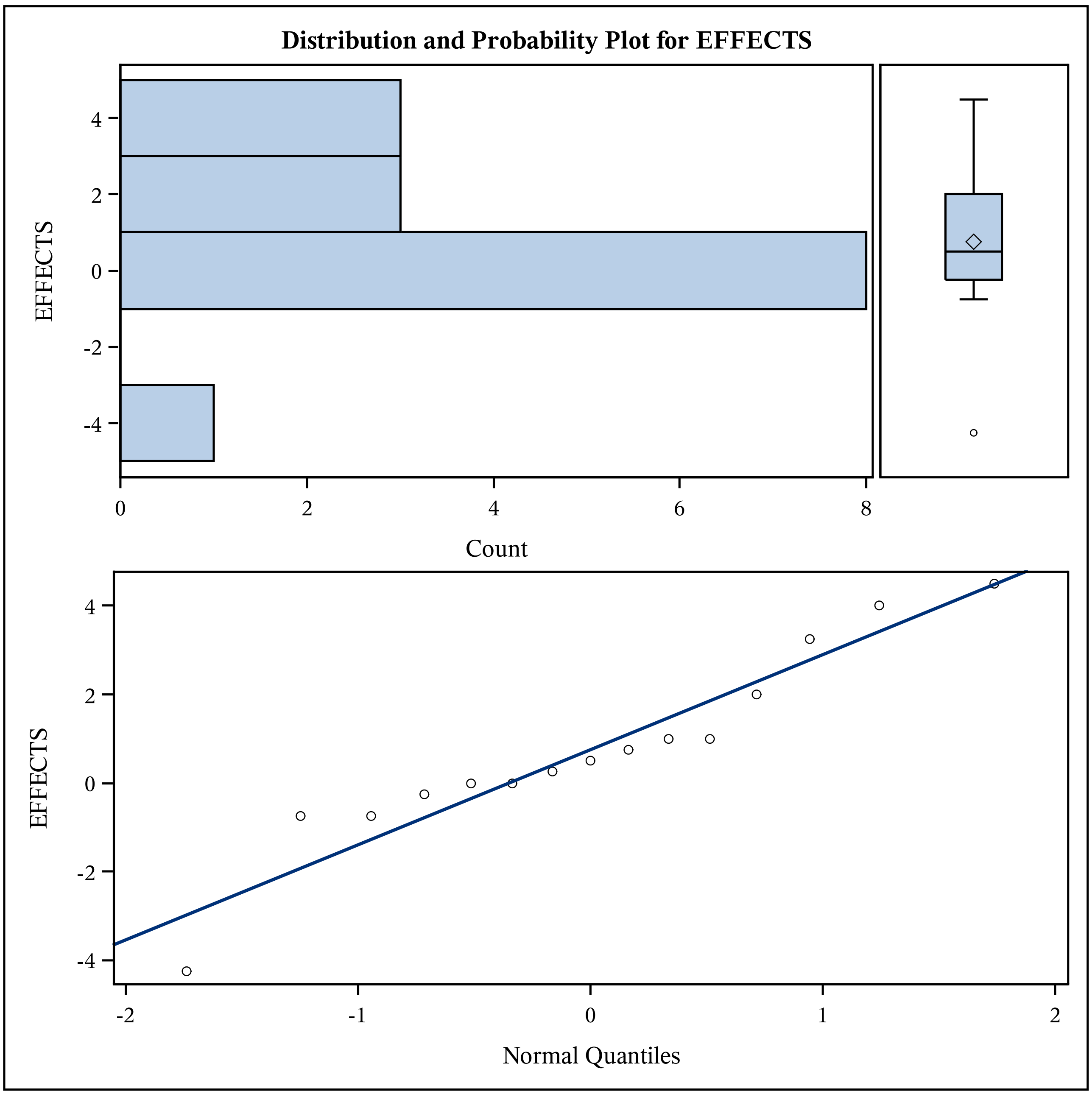
Make a NPP of these estimates

```
DM 'LOG; CLEAR; OUT; CLEAR;';
ODS LISTING;
* ODS PRINTER PDF file='C:\COURSES\ST578\SAS\TWO4.PDF';
OPTIONS PS=54 LS=78 NODATE NONUMBER;

DATA IN;
  DO TEMP     = -1 TO 1 BY 2;
  DO PRESSURE = -1 TO 1 BY 2;
  DO CONC     = -1 TO 1 BY 2;
  DO TIME     = -1 TO 1 BY 2;
     INPUT YIELD @@;  OUTPUT;
  END; END; END; END;
 LINES;
 12 18 13 16 17 15 20 15 10 25 13 24 19 21 17 23
 ;
**********************************************************;
*** PART I:  DETERMINE THE ESTIMATES OF THE 15 EFFECTS ***;
**********************************************************;

PROC GLM DATA=IN;
 CLASS TIME CONC PRESSURE TEMP;
 MODEL YIELD = TIME|CONC|PRESSURE|TEMP / SS3;

 ESTIMATE 'TIME'     TIME     -1 1;
 ESTIMATE 'CONC'     CONC     -1 1;
 ESTIMATE 'PRESSURE' PRESSURE -1 1;
 ESTIMATE 'TEMP'     TEMP     -1 1;

 ESTIMATE 'TIME*CONC' TIME*CONC     1 -1 -1 1 / DIVISOR=2;
 ESTIMATE 'TIME*PRES' TIME*PRESSURE 1 -1 -1 1 / DIVISOR=2;
 ESTIMATE 'TIME*TEMP' TIME*TEMP     1 -1 -1 1 / DIVISOR=2;
 ESTIMATE 'CONC*PRES' CONC*PRESSURE 1 -1 -1 1 / DIVISOR=2;
 ESTIMATE 'CONC*TEMP' CONC*TEMP     1 -1 -1 1 / DIVISOR=2;
 ESTIMATE 'PRES*TEMP' PRESSURE*TEMP 1 -1 -1 1 / DIVISOR=2;

 ESTIMATE 'TIME*C*P' TIME*CONC*PRESSURE -1 1 1 -1 1 -1 -1 1 / DIVISOR=4;
 ESTIMATE 'TIME*C*T' TIME*CONC*TEMP     -1 1 1 -1 1 -1 -1 1 / DIVISOR=4;
 ESTIMATE 'TIME*P*T' TIME*PRESSURE*TEMP -1 1 1 -1 1 -1 -1 1 / DIVISOR=4;
 ESTIMATE 'C*P*TEMP' CONC*PRESSURE*TEMP -1 1 1 -1 1 -1 -1 1 / DIVISOR=4;

 ESTIMATE 'T*C*P*T' TIME*CONC*PRESSURE*TEMP
           1 -1 -1 1 -1 1 1 -1   -1 1 1 -1 1 -1 -1 1 / DIVISOR=8;
TITLE 'A 2**4 DESIGN -- ESTIMATION OF EFFECTS';
```

```
***********************************************************************;
*** PART II:   MAKE A NORMAL PROBABILITY PLOT OF THE ESTIMATED EFFECTS ***;
***********************************************************************;

DATA FX; INPUT EFFECTS @@; LINES;
 4.5 0.5 2 3.25 -0.75 -4.25 4 0.25 0 0 1 0.75 -0.25 -0.75 1
 ;
PROC UNIVARIATE DATA=FX PLOTS;
     VAR EFFECTS;
TITLE 'A 2**4 DESIGN -- NORMAL PROBABILITY PLOT OF EFFECTS';
```

A 2**4 DESIGN -- NORMAL PROBABILITY PLOT OF EFFECTS

The UNIVARIATE Procedure



Analysis I: Pooling high order interactions

- After pooling all 3-factor and 4-factor interaction, we have 5 df for the MS_E.
- The ANOVA indicates significant A, C, AC, D, and AD effects. These match the highlighted points on the normal probability plot of effects.

```
****************************************************************;
*** PART III:  RUN ANOVA WITH POOLED HIGHER ORDER INTERACTIONS ***;
****************************************************************;

PROC GLM DATA=IN;
     CLASS TIME CONC PRESSURE TEMP;
     MODEL YIELD = TIME|CONC|PRESSURE|TEMP@2 / SS3;
TITLE 'A 2**4 DESIGN -- POOLING HIGHER ORDER INTERACTIONS';
```

A 2**4 DESIGN -- POOLING HIGHER ORDER INTERACTIONS

Dependent Variable: YIELD

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	279.00000	27.90000	10.94	0.0083
Error	5	12.75000	2.55000		
Corrected Total	15	291.75000			

R-Square	C.V.	Root MSE	YIELD Mean
0.956298	9.190630	1.5969	17.37500

Source	DF	Type III SS	Mean Square	F Value	Pr > F	
TIME	1	81.000000	81.000000	31.76	0.0024	→ A
CONC	1	1.000000	1.000000	0.39	B – 0.5586	
TIME*CONC	1	2.250000	2.250000	0.88	AB – 0.3907	
PRESSURE	1	16.000000	16.000000	6.27	0.0542	→ C
TIME*PRESSURE	1	72.250000	72.250000	28.33	0.0031	→ AC
CONC*PRESSURE	1	0.250000	0.250000	0.10	B-C- 0.7668	
TEMP	1	42.250000	42.250000	16.57	0.0096	→ D
TIME*TEMP	1	64.000000	64.000000	25.10	0.0041	→ AD
CONC*TEMP	1	0.000000	0.000000	0.00	BD 1.0000	
PRESSURE*TEMP	1	0.000000	0.000000	0.00	CD 1.0000	

Analysis II: Pooling terms involving factor B = concentration (CONC)

- After pooling all terms involving CONC, we have 8 df for the MS_E.
- The ANOVA indicates significant A, C, AC, D, and AD effects. These match the highlighted points on the normal probability plot of effects.
- After factor B is removed, we still retain balance and orthogonality. We now have a 2^3 design with $n = 2$ replicates for each combination of factor levels for A, C, and D.

```
**************************************************************;
*** RUN ANOVA WITH CONCENTRATION REMOVED FROM THE ANALYSIS ***;
**************************************************************;

PROC GLM DATA=IN;
     CLASS TIME PRESSURE TEMP;
     MODEL YIELD = TIME|PRESSURE|TEMP / SS3;
TITLE 'ANOVA WITH CONCENTRATION REMOVED FROM THE ANALYSIS';

RUN;
```

ANOVA WITH CONCENTRATION REMOVED FROM THE ANALYSIS

Dependent Variable: YIELD

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	275.75000	39.39286	19.70	0.0002
Error	8	16.00000	2.00000		
Corrected Total	15	291.75000			

R-Square	C.V.	Root MSE	YIELD Mean
0.945159	8.139359	1.4142	17.37500

Source	DF	Type III SS	Mean Square	F Value	Pr > F	
TIME	1	81.000000	81.000000	40.50	0.0002	A
PRESSURE	1	16.000000	16.000000	8.00	0.0222	C
TIME*PRESSURE	1	72.250000	72.250000	36.13	0.0003	A·C
TEMP	1	42.250000	42.250000	21.12	0.0018	D
TIME*TEMP	1	64.000000	64.000000	32.00	0.0005	A·D
PRESSURE*TEMP	1	0.000000	0.000000	0.00	1.0000	
TIME*PRESSURE*TEMP	1	0.250000	0.250000	0.13	0.7328	