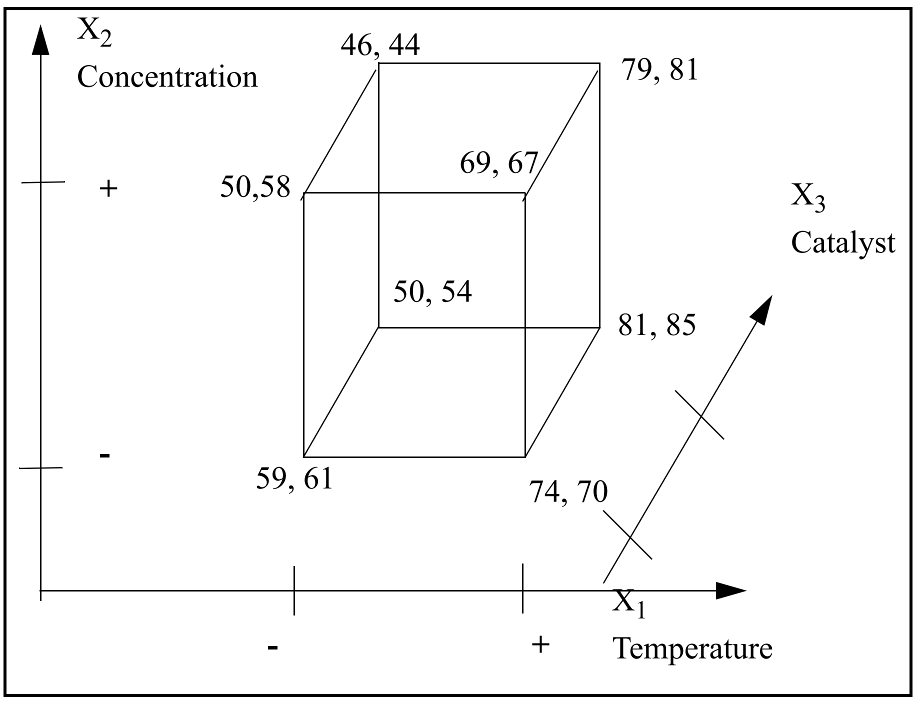
10.0 REPLICATED FULL FACTORIAL DESIGN

(Updated Spring, 2001)

Pilot Plant Example (2^3), resp - Chemical Yield%

	Lo(-1)	Hi(+1)
Temperature	160°	180° C
Concentration	10%	40%
Catalyst	A	B

Test#	Temp	Conc	Catalyst	Yield		
	x_1	x_2	x_3	y_{i1}	y_{i2}	$\bar{y}$
1	-1	-1	-1	59	61	60
2	+1	-1	-1	74	70	72
3	-1	+1	-1	50	58	54
4	+1	+1	-1	69	67	68
5	-1	-1	+1	50	54	52
6	+1	-1	+1	81	85	83
7	-1	+1	+1	46	44	45
8	+1	+1	+1	79	81	80



Standard Order

- 1st variable column (-1,+1,-1...)
  Alternates every 2^0 value
- 2nd variable column (-1,-1,+,+,-,-,+,+)
  Alternates every 2^1 values
- 3rd variable column (-,-,-,-,+,+,+,+,-,-,-,-,+,+,+,+,...)
  Alternates every 2^2 values

- k^{th} variable - alternates every 2^{k-1} values

Underlying Model

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3$$
$$+ b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3$$
$$+ b_{123}x_1x_2x_3 + \varepsilon$$

Avg. of all responses at high level of x_1

$$= \frac{74 + 70 + 69 + 67 + 81 + 85 + 79 + 81}{8} = 75.75$$

Avg. of all responses at low level for x_1

$$= \frac{59 + 61 + 50 + 58 + 50 + 54 + 46 + 44}{8} = 52.75$$

E_1 = 75.75 - 52.75=23.0   Geometrical interpret

$$E_1 = \frac{-59 - 61 + 74 + 70 - 50 \ldots + 85 - 46 - 44 + 79 + 81}{\text{Number of } '' + '' \text{Signs} = 8}$$

E_1 = 23.0

$$E_1 = \frac{-60 + 72 - 54 + 68 - 52 + 83 - 45 + 80}{4} = 23.0$$

$\bar{y}$'s (We will work with these)

Calculation Matrix

Test	I	1	2	3	12	13	23	123	$\bar{Y}_i$
1	+	-	-	-	+	+	+	-	60
2	+	+	-	-	-	-	+	+	72
3	+	-	+	-	-	+	-	+	54
4	+	+	+	-	+	-	-	-	68
5	+	-	-	+	+	-	-	+	52
6	+	+	-	+	-	+	-	-	83
7	+	-	+	+	-	-	+	-	45
8	+	+	+	+	+	+	+	+	80

$$\text{Avg} = \frac{60 + 72 + 54 + 68 + 52 + 83 + 45 + 80}{8} = 64.25$$

$$E_1 = \frac{-60 + 72 - 54 + 68 - 52 + 83 - 45 + 80}{4} = 23.0$$

$$E_{12} = \frac{+60 - 72 - 54 + 68 + 52 - 83 - 45 + 80}{4} = 1.5$$

$$E_{123} = \frac{-60 + 72 + 54 - 68 + 52 - 83 - 45 + 80}{4} = 0.5$$

Summary

Avg = 64.25	E_{12} =1.5
E_1 = 23.0	E_{13} =10.0
E_2 = -5.0	E_{23} = 0.0
E_3 = 1.5	E_{123} =0.5

- Which of these are important?
- Which of the effects are distinguishable from the noise in the experimental environment?

Lets say we run experiment numerous times. Sample effects, E's, are distributed normally.

E = Linear Combination of y's

We will look at three ways to figure out the important effects

- Replication of experiment
- Normal Prob. Plot
- Make an assumption about higher - order effects being negligible.

Replication -> Important Effects

Our experiment is replicated, calculate a sample variance for each test.

Test	y_{i1}	y_{i2}	$\bar{y}$	S_y^2
1	59	61	60	2
2	74	70	72	8
3	50	58	54	32
4	69	67	68	2
5	50	54	52	8
6	81	85	83	8
7	46	44	45	2
8	79	81	80	2

Let's assume σ_y^2 is constant for all the tests.

Calculate pooled sample variance estimate of σ_y^2

$$S_P^2 = \frac{\Sigma s_P^2}{8} = \frac{2+8+32+2+8+8+2+2}{8} = 8$$

<u>Pilot Plant Example: Repl. 2^3 Factorial</u>

Test	s_p^2	x_1	x_2	x_3	y_{i1}	y_{i2}	$\bar{y}$
1	2	-	-	-	59	61	60
2	8	+	-	-	74	70	72
3	32	-	+	-	50	58	54
4	2	+	+	-	69	67	68
5	8	-	-	+	50	54	52
6	8	+	-	+	81	85	83
7	2	-	+	+	46	44	45
8	2	+	+	+	79	81	80

Calculated effects: E_i is an estimate of μ_E

Avg = 64.25	E_{12} =1.5
E_1 = 23.0	E_{13} =10.0
E_2 = -5.0	E_{23} = 0.0
E_3 = 1.5	E_{123} =0.5

$$S_P^2 = \frac{\sum_{i=1}^{m} s_i^2}{m}$$

where $m=2^k$,
n is the number of trials for each test condition,
N is total number of tests, and N=mn

Under H_o: $(\mu_{E_1} = \mu_{E_2} = \ldots = 0)$

The E's have arisen from a normal distribution centered at 0.

We may, based on each calculated effect, develop a confidence interval for the true mean effect

$E_i \pm t\, s_{eff}$ -> How do we estimate σ_{eff}

First let's develop a confidence interval for the true average

$$Avg = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} y_{ij}}{mn} = 64.25$$

in our case

$$\text{Var (Avg)} = \text{Var}\left[\frac{y + y + \ldots + y}{N}\right] = \frac{1}{N^2} Var[y + y + \ldots + y]$$

$$= \frac{N\sigma_y^2}{N^2} = \frac{\sigma_y^2}{N} \text{ -> as expected}$$

$$s_{avg}^2 = \frac{S_p^2}{N} = \frac{8}{16} = 0.5$$

$$s_{avg} = 0.707$$

$100(1-\alpha)\%$ Confidence interval for average (95% C.I.) is

$$Avg \pm t_{\nu, 1-\frac{\alpha}{2}} s_{avg} = 64.25 \pm t_{8, 0.975} s_{avg} = 64.25 \pm (2.306)(0.707)$$

Since 0 is not on C.I., reject H_o that $\mu_{avg} = 0$

Confidence Intervals for Other Effects

We calculate effects, generally speaking by:

$$E = \bar{y}(+) - \bar{y}(-)$$

Each based on N/2 observations (8 for our example)

$$Var[E] = Var[\bar{y}(+)] + Var[\bar{y}(-)]$$

$$Var[\bar{y}(+)] = Var\left[\frac{y + y + \ldots + y}{N/2}\right]$$

$$= \frac{4}{N^2} Var[y+y+...+y]$$

$$= \frac{4}{N^2}\left(\frac{N}{2}\sigma_y^2\right) = \frac{2}{N}\sigma_y^2$$

$$Var[E] = \frac{2}{N}\sigma_y^2 + \frac{2}{N}\sigma_y^2$$

$$\sigma_{eff}^2 = \frac{4}{N}\sigma_y^2$$

For our example,

$$s_{eff}^2 = \frac{4}{N}s_P^2 = \frac{4}{16}.8 = 2$$

Total number of trials = 16,

$$s_{eff} = 1.414$$

Sometimes called the Std. error of an effect
$100(1-\alpha)\%$ confidence interval

$$Avg \pm t_{\nu, 1-\frac{\alpha}{2}} s_{avg}$$

95% confidence interval is

$$E_i \pm (2.306)(1.414)$$

$$E_i \pm 3.26$$

95% confidence interval for each effect

$E_1 = 23.00 \pm 3.26$* $\quad$ $E_{12} = 1.50 \pm 3.26$

$E_2 = -5.00 \pm 3.26$* $\quad$ $E_{13} = 10.00 \pm 3.26$*

$E_3 = 1.50 \pm 3.26$ $\quad$ $E_{23} = 0 \pm 3.26$

$E_{123} = 0.5 \pm 3.26$

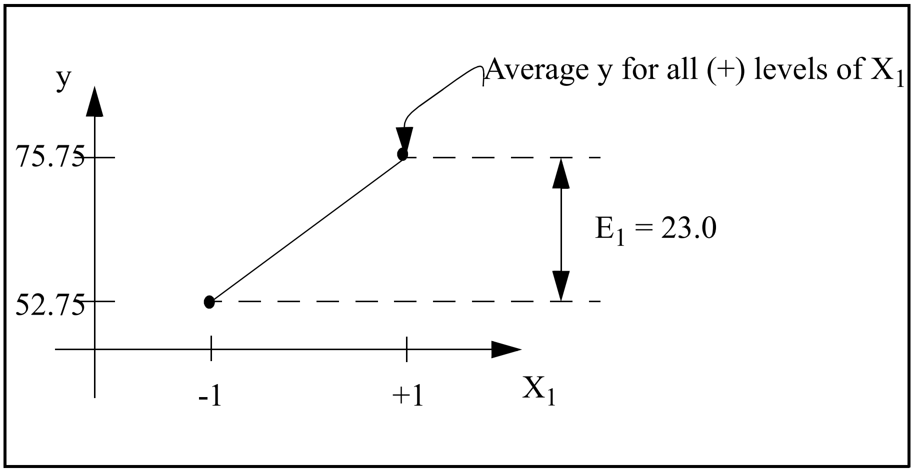
0 does not lie on C.I. for E_1, E_2, & E_{13}.
The calculated (sample) effects E_1, E_2, & E_{13} have arisen from a normal distribution not centered at 0.

Interpreting Significant Effects

- $E_1 = 23.0$

Collapse or average the design in X_2 and X_3 directions

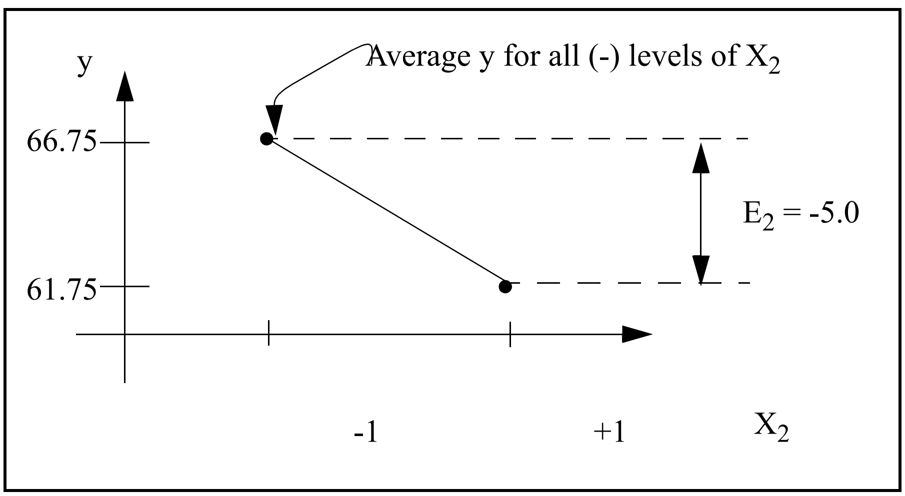
.



- $E_2 = -5.0$

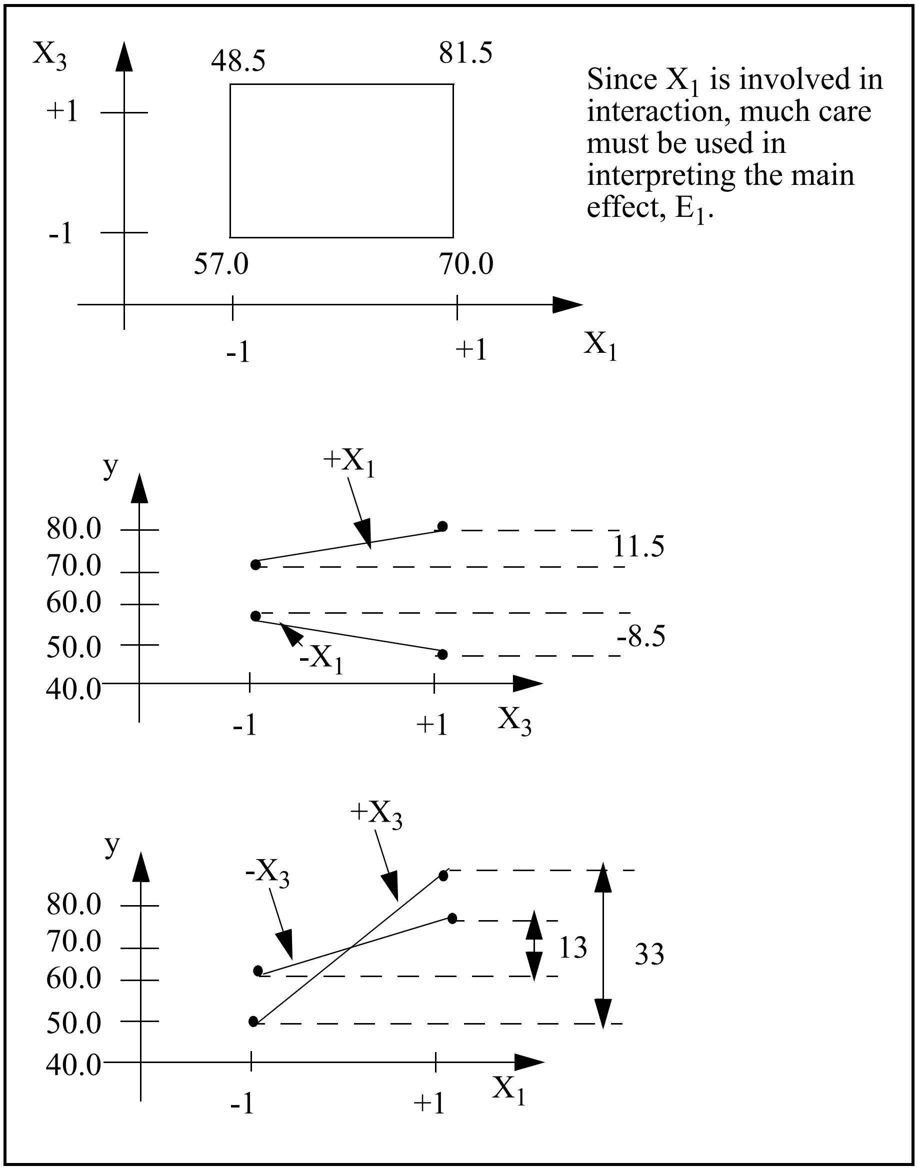
  Collapse or average the design in X_1 and X_3 directions

.



- $E_{13} = 10.0$

  Collapse or average the design in X_2 direction.



Mathematical Model

For 2^3 factorial design pilot plant example we tacitly assume that response can be characterized as:

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{23} x_2 x_3 + b_{123} x_1 x_2 x_3 + \varepsilon$$

We ran tests and fit this equation to the data:

$$\hat{y} = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \hat{b}_3 x_3 + \hat{b}_{12} x_1 x_2 + \hat{b}_{13} x_1 x_3 + \hat{b}_{23} x_2 x_3 + \hat{b}_{123} x_1 x_2 x_3$$

We now know that only $\hat{b}_0$, $\hat{b}_1$, $\hat{b}_2$ & $\hat{b}_{13}$ are important.
The fitted model becomes:

$$\hat{y} = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \hat{b}_{13} x_1 x_3$$

$$\hat{y} = Avg + \left(\frac{E_1}{2}\right)x_1 + \left(\frac{E_2}{2}\right)x_2 + \left(\frac{E_{13}}{2}\right)x_1 x_3$$

$$\hat{y} = 64.25 + 11.5\, x_1 - 2.5\, x_2 + 5 x_1\, x_3$$

For each unique combination of x_1, x_2, & x_3 a predicted response may be calculated.

$$\hat{y} = 64.25 + 11.5\,(-1) - 2.5\,(-1) + 5(-1)\,(-1)$$

Test#	x_1	x_2	x_3	y_{ij}		$\bar{y}_i$	$\hat{y}_i$	$e_{ij} =$	$y_{ij} - \hat{y}$
1	-	-	-	59	61	60	60.25	-1.25	0.75
2	+	-	-	74	70	72	73.25	0.75	-3.25
3	-	+	-	50	58	54	55.25	-5.25	2.75
4	+	+	-	69	67	68	68.25	0.75	-1.25
5	-	-	+	50	54	52	50.25	-0.25	3.75
6	+	-	+	81	85	83	83.25	-2.25	1.75
7	-	+	+	46	44	45	45.25	0.75	-1.25
8	+	+	+	79	81	80	78.25	0.75	2.75

We must examine the model residuals to search for model inadequacies.

Plots:
- e_{ij} vs. run order
- e vs. $\bar{y}$
- e_{ij} vs. x_1, x_2,...
- $(y_{ij} - \bar{y}_i)$ vs. $\bar{y}_i$

Using the Model for Process Improvement

$$\hat{y} = 64.25 + 11.5\, x_1 - 2.5\, x_2 + 5 x_1\, x_3$$

- What values of x_1, x_2, & x_3 maximize predicted yield?
- How does the response behave as a function of x_1, x_2, & x_3?
- What is the shape of the response surface?

$$x_1 = \frac{Temp - \frac{1}{2}(Temp_{HI} + Temp_{LO})}{\frac{1}{2}(Temp_{HI} - Temp_{LO})}$$

$$x_1 = \frac{Temp - MidTemp}{UnitChangeInTemp} = \frac{Temp - 170}{10}$$

• When Temp = 165

$$x_1 = \frac{165 - 170}{10} = -0.5$$

• When x_1 = -0.2

$$-0.2 = \frac{Temp - 170}{10} \text{ -> Temp} = 170 + (-.2)(10) = 168$$

• Only can use this transformation for quantitative variables.

Recall that:

		Low (-1)	High(+1)
x_1	Temp	160	180
x_2	Conc.	20	40
x_3	Catalyst	A	B

Based on other criteria, catalyst set to A, i.e., $x_3 = -1$

So, $\hat{y}$ = 64.25 + 11.5 x_1 - 2.5 x_2 +5x_1 (-1)

or $\hat{y}$ = 64.25 + 6.5 x_1 - 2.5 x_2

The presence of interactions warp the plane.

<u>Remember, 3 ways to figure out the significant effects:</u>

1. Use replication to get s_p^2 (estimate of σ_y^2) $\rightarrow s_{eff}^2$ (estimate of σ_{eff}^2)
2. Assume that higher order interactions are negligible $\rightarrow s_{eff}^2$ (estimate of σ_{eff}^2)
3. Normal probability plot.

Higher Order Interactions Negligible

Study surface finish (Ra value) produced by a machining operation.

Variable	Low (-1)	High(+1)
1-Speed (ft/min)	1000	2000
2 - Feed (in./rev)	.005	.015
3 - Depth of Cut (in.)	.050	.100
4 - Nose Radius (in.)	1/64	3/64
5 - SCEA	15^o	30^o

Calculated Effects - Effect Estimates

Avg = 171.773
E_1 = -2.087
E_2 = 274.496
E_3 = -1.752
E_4 = -171.752
E_5 = 0.042
E_{12} = -2.018
E_{13} = 3.634
E_{14} = -0.008
E_{15} = -1.242
E_{23} = -1.953
E_{24} = -137.208
E_{25} = -0.077
E_{34} = 0.389
E_{35} = 2.041
E_{45} = -0.087

E_{123} = 2.658
E_{124} = -0.134
E_{125} = -0.271
E_{134} = -3.672
E_{135} = -3.968
E_{145} = -1.078
E_{234} = 1.033
E_{235} = 1.447
E_{245} = 0.374
E_{345} = -0.648
E_{1234} =-3.053
E_{1235} = -4.407
E_{1245} = -1.687
E_{1345} = 3.501
E_{2345} = -0.292
E_{12345} = 4.212

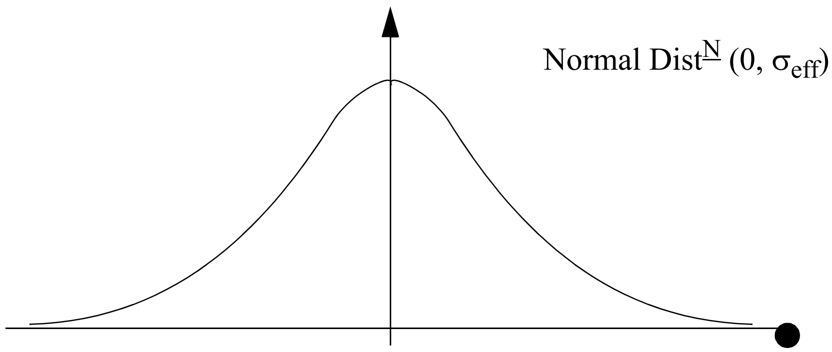
Test Table

Test	1	2	3	4	5	R_a
1	-	-	-	-	-	51.09
2	+	-	-	-	-	52.03
3	-	+	-	-	-	485.6
4	+	+	-	-	-	451.28
5	-	-	+	-	-	50.54
6	+	-	+	-	-	52.37
7	-	+	+	-	-	446.94
8	+	+	+	-	-	470.83
9	-	-	-	+	-	17.87
10	+	-	-	+	-	17.66
11	-	+	-	+	-	156.73
12	+	+	-	+	-	156.73
13	-	-	+	+	-	16.56
14	+	-	+	+	-	17.61
15	-	+	+	+	-	151.87
16	+	+	+	+	-	152.13
17	-	-	-	-	+	53.45
18	+	-	-	-	+	48.93
19	-	+	-	-	+	463.51
20	+	+	-	-	+	463.87
21	-	-	+	-	+	52.50
22	+	-	+	-	+	53.47
23	-	+	+	-	+	465.88
24	+	+	+	-	+	460.10
25	-	-	-	+	+	17.38
26	+	-	-	+	+	16.99
27	-	+	-	+	+	158.25
28	+	+	-	+	+	150.82
29	-	-	+	+	+	17.09
30	+	-	+	+	+	16.87
31	-	+	+	+	+	159.61
32	+	+	+	+	+	149.99

Comments

- As we go from $2^3 \rightarrow 2^4 \rightarrow 2^5 \rightarrow 2^6$ generally the proportion of sig. effects decreases
- Response surfaces generally fairly smooth - absence of higher order interactions
- Rarely find 3 - factor interactions significant, even rarer to find 4 - factor interactions significant even rarer to find 5 - factor interactions significant and so on.

Let’s assume that higher order interactions are negligible. Specifically, let’s assume that 3 - factor and higher order interactions are negligible. The calculated effect estimates are then non-zero only because of experimental error.



Calculated Effects Value

(True effect value=0 for high order interactions)

Under our assumption, all calculated 3-factor interactions come from a normal dist$^{\underline{n}}$ with a standard deviation of σ_{eff} centered at 0. Use the calculated effects to estimate σ_{eff}.

$$s_{eff}^2 = \sum_{\text{All Higher Order Effects}} \frac{(E_i)^2}{\text{Number of 3rd and higher order}}$$

In above equation, the number of higher order effects equals the number of degrees of freedom. This is because that we have already assumed zero mean for the normal distribution.

If we take 3rd and higher order interactions,

$$s_{eff}^2 = \frac{(2.658)^2 + (-.134)^2 + \ldots + (4.212)^2}{16} = 6.43353$$

$s_{eff} = 2.536$ is an estimate of σ_{eff}.

Construct $100\ (1-\alpha)\%$ Confidence Interval for each effect estimate

$$E_i \pm t_{\nu,\,1-\frac{\alpha}{2}} s_{eff}$$

where ν is the number of 3 - factor & higher order interactions used to calculate s_{eff}. For our example, $\nu = 16$, $\alpha = 0.05$,

The C.I. is

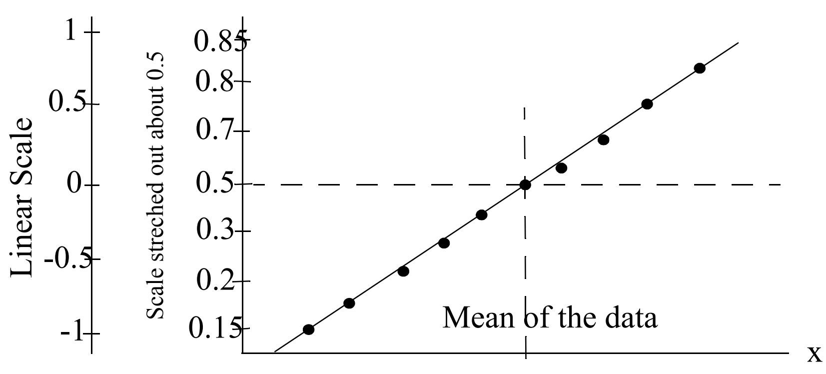
$$E_i \pm (2.12)(2.536) = E_i \pm 5.377$$

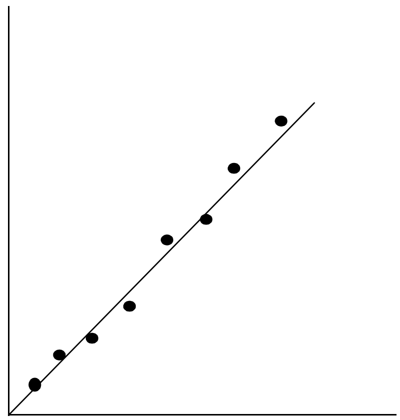
We can conclude that the average, E_2, E_4, & E_{24} are significant. One final way to figure out which effects are significant is the normal probability plot.

Normal Probability Plot

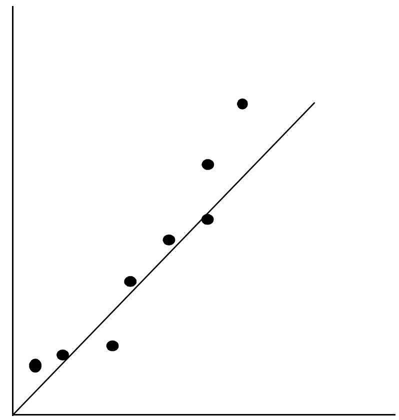
Draw 100 values from a normal dist$^{\underline{n}}$, and plot the CDF for it. Draw another 100 values from an unknown dist$^{\underline{n}}$ and plot the CDF for it too. Is the unknown dist$^{\underline{n}}$ normal? Does the "S" shaped appearance in the data match the S shapes of the normal CDF?

Let's change the vertical scale on the CDF plot so that the normal CDF becomes a straight line





Normal



Not Normal

If the sampled data from the unknown dist$^{\underline{n}}$ appears to be well described by a line on such a plot, then the unknown dist$^{\underline{n}}$ may be reasonably approximated as normal.

How can we use a normal probability plot to help us decide which effects are important?

- Calculate the effects
- Rank the effects from smallest to largest (don't include the average)
- Calculate the cumulative prob. associated with each effect.

$$P_i = \frac{100\left(i - \frac{1}{2}\right)}{N}$$

- Plot the points (E_i, P_i) on Normal probability paper (Home-made or purchased).