Are the estimated model parameters significant?

If we started the whole experiment from the begining would we get the same results?

$$y = Y + \varepsilon$$

y varies around Y

If we analysed the new results would we get the same estimates for the model parmeters?

b_j is a random variable, its value differs from zero, even if β_j=0

Are the estimated b parameters significantly differ from zero?

$$t = \frac{b_j - \beta_j}{s_{b_j}}$$

$$s_{b_j}^2 = \frac{s_y^2}{\sum_i x_{ji}^2} = \frac{s_y^2}{N}$$

Null hypothesis: $H_0 : \beta_j = 0$

If the null hypothesis is true the ratio (b_j/s_{bj}) is t-distributed.

$$P\left(-t_{\alpha/2} < b_j / s_{b_j} \le t_{\alpha/2}\right) = 1 - \alpha$$

The null hypothesis is rejected if: $\left|b_j\right| > s_{b_j} t_{\alpha/2}$

Where do we get s_y^2 ?

They made repetitions in the centrum (where the level of each factor is 0):

i	Natural unit		Transformed factors		*y*
	z_1	z_2	x_1	x_2	
5	0.25	27.5	0	0	50
6	0.25	27.5	0	0	50
7	0.25	27.5	0	0	51

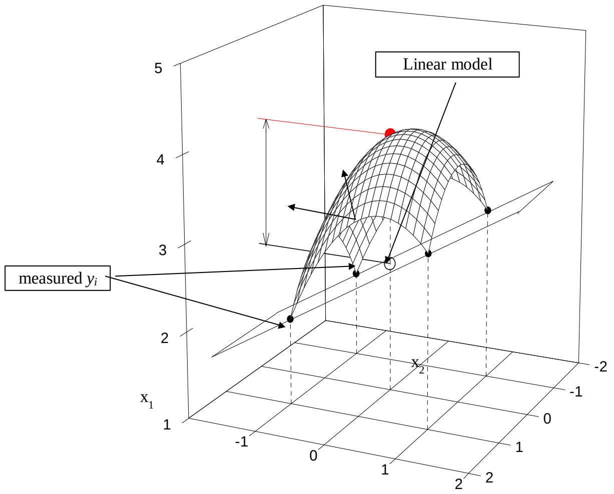
$$\bar{y}_1^0 = \frac{\sum_{m=1}^{3} y_{1m}^0}{3} = 50.33 \qquad s_{y_1 0}^2 = \frac{\sum_{m=1}^{3}\left(y_{1m}^0 - \bar{y}_1^0\right)^2}{2} = 0.333$$

$$s_{b_j}^2 = \frac{s_y^2}{\sum_i x_{ji}^2} = \frac{s_y^2}{N} = \frac{0.333}{4} = 0.0833 \qquad s_{b_j} = 0.289$$

$$P\left(-t_{\alpha/2} < b_j \big/ s_{b_j} < t_{\alpha/2}\right) = 1 - \alpha \qquad \text{if } b_j = 0$$

Significant if $\left|b_j\right| > s_{b_j} t_{\alpha/2}$

$t_{0.05/2} = 4.3$ $\qquad s_{b_j} t_{\alpha/2} = 0.289 \cdot 4.3 = 1.243$



Checking goodness of fit for the linear model (curvature check)

$E\left(\bar{y}^0\right) = Y^0$ (centre point)

H_0: $E\left(b_0\right) = E\left(\bar{y}^0\right) = Y^0$

H_1: $E\left(b_0\right) \neq E\left(\bar{y}^0\right) = Y^0$

$t_0 = \dfrac{d}{s_d}$ $d = \bar{y}^0 - b_0$ $s_d^2 = s_y^2\left(\dfrac{1}{k_c} + \dfrac{1}{N}\right)$

$\nu = N - l + k_c - 1$

$$d = 50.33 - 50 = 0.33$$

$$s_d^2 = 0.333\left(\frac{1}{4} + \frac{1}{3}\right) = 0.1943 \qquad s_d = 0.441$$

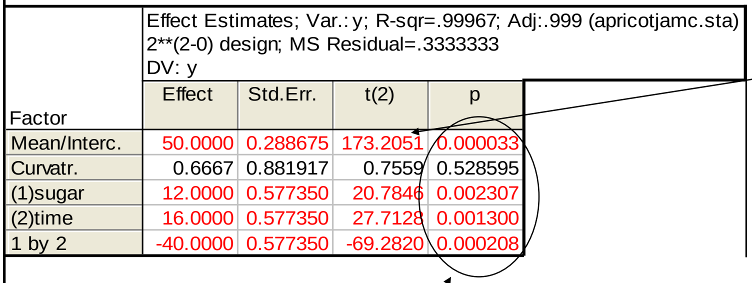
$$t_0 = \frac{0.33}{0.441} = 0.748$$

$$\nu = N - l + k_c - 1 = 4 - 4 + 3 - 1 = 2$$

$$t_0 < t_{0.05/2}(2) = 4.3$$

The null hypothesis is accepted
(no quadratic term is required)

Quick >Summary: Effect estimates



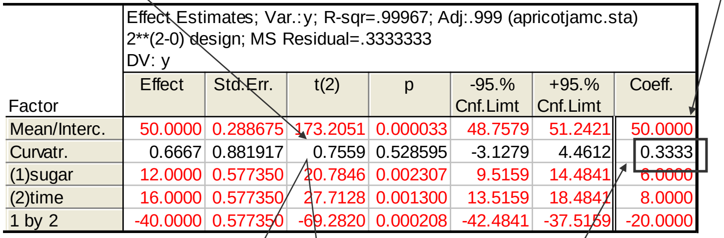
Effect Estimates; Var.:y; R-sqr=.99967; Adj:.999 (apricotjamc.sta)
2**(2-0) design; MS Residual=.3333333
DV: y

Factor	Effect	Std.Err.	t(2)	p
Mean/Interc.	50.0000	0.288675	173.2051	0.000033
Curvatr.	0.6667	0.881917	0.7559	0.528595
(1)sugar	12.0000	0.577350	20.7846	0.002307
(2)time	16.0000	0.577350	27.7128	0.001300
1 by 2	-40.0000	0.577350	-69.2820	0.000208

$$t(2) = \frac{\text{Effect}}{\text{Std.Err}}$$

Probability of obtaining t(2) values of this size or more extreme if the true effect is zero (null hypothesis) *t*-test for curvature

value calculated without centre points



Effect Estimates; Var.:y; R-sqr=.99967; Adj:.999 (apricotjamc.sta)
2**(2-0) design; MS Residual=.3333333
DV: y

Factor	Effect	Std.Err.	t(2)	p	-95.% Cnf.Limt	+95.% Cnf.Limt	Coeff.
Mean/Interc.	50.0000	0.288675	173.2051	0.000033	48.7579	51.2421	50.0000
Curvatr.	0.6667	0.881917	0.7559	0.528595	-3.1279	4.4612	0.3333
(1)sugar	12.0000	0.577350	20.7846	0.002307	9.5159	14.4841	6.0000
(2)time	16.0000	0.577350	27.7128	0.001300	13.5159	18.4841	8.0000
1 by 2	-40.0000	0.577350	-69.2820	0.000208	-42.4841	-37.5159	-20.0000

$$\mathrm{t}(2) = \frac{0.6667}{0.881917} = \frac{0.3333}{0.440959} = 0.7559$$

deviation of the mean of centre values from the Mean

How to estimate the σ^2 variance $\left(s_y^2\right)$

k repetitions at all points of the design

Number of runs: $N = k{\cdot}2^p$

One may check the σ^2=const condition

Adequacy of the linear model may not be checked

k_c repetitions at the centre of the design

Number of runs: $N = 2^p + kc << k{\cdot}2^p$

One may not check the σ^2=const condition

Adequacy of the linear model may be checked.

k repetitions at all points of the design, k_c at the centre

Number of runs : $N = k{\cdot}2^p+k_c$

One may not check the *s*2=const condition

Adequacy of the linear model may be checked

More sensitive significance tests due to larger df

What happens if only one point is repeated?

i	x_0	x_1	x_2	x_1x_2
1	+	–	–	+
2	+	+	–	–
3	+	–	+	–
4	+	+	+	+
5	+	+	+	+

$$\sum_i x_{ji}x_{ki} = 0, \text{ if } \quad j \neq k$$

orthogonality?

Example

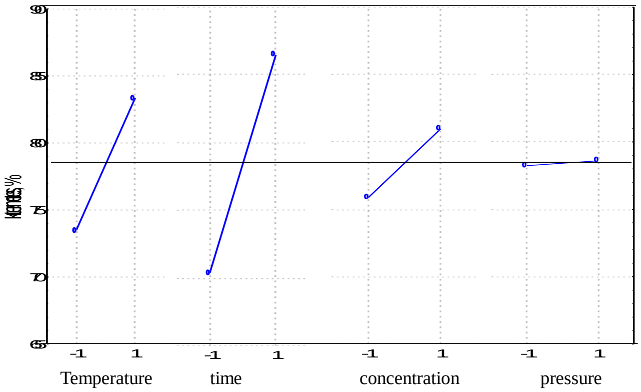
The yield (%) of a chemical reactor is investigated as a function of 4 factors

z_1	temperature 40 and 60 ^{o}C
z_2	reaction time 10 and 20 min
z_3	concn. of the initial component 45 and 65 %
z_4	pressure 2 and 6 bar

4fact.xls

Factors	z_1	z_2	z_3	z_4
center point z_j^0	50	15	55	4
variation interval Δz_j	10	5	10	2
upper level z_j^{max} (+)	60	20	65	6
lower level z_j^{min} (–)	40	10	45	2

	In natural units				Coded levels					y
i	z_1	z_2	z_3	z_4	x_0	x_1	x_2	x_3	x_4	%
1	40	10	45	2	+	–	–	–	–	60.4
2	60	10	45	2	+	+	–	–	–	75.9
3	40	20	45	2	+	–	+	–	–	79.8
4	60	20	45	2	+	+	+	–	–	86.0
5	40	10	65	2	+	–	–	+	–	64.9
6	60	10	65	2	+	+	–	+	–	80.9
7	40	20	65	2	+	–	+	+	–	86.4
8	60	20	65	2	+	+	+	+	–	91.6
9	40	10	45	6	+	–	–	–	+	59.6
10	60	10	45	6	+	+	–	–	+	77.0
11	40	20	45	6	+	–	+	–	+	83.1
12	60	20	45	6	+	+	+	–	+	85.0
13	40	10	65	6	+	–	–	+	+	65.0
14	60	10	65	6	+	+	–	+	+	79.3
15	40	20	65	6	+	–	+	+	+	88.7
16	60	20	65	6	+	+	+	+	+	91.1

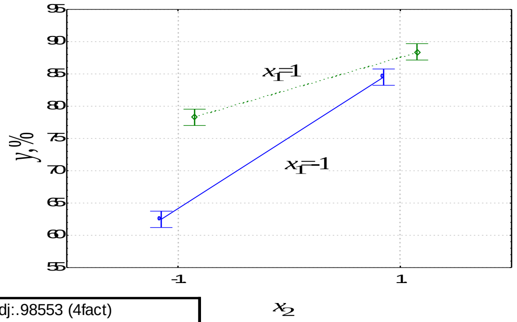


$$\hat{Y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 +$$

$$+ b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{14} x_1 x_4 + b_{23} x_2 x_3 + b_{24} x_2 x_4 + b_{34} x_3 x_4 +$$

$$+ b_{123} x_1 x_2 x_3 + b_{124} x_1 x_2 x_4 + b_{134} x_1 x_3 x_4 + b_{234} x_2 x_3 x_4 + b_{1234} x_1 x_2 x_3 x_4$$

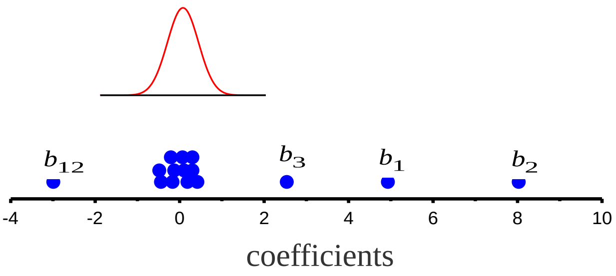
i	x_0	x_1	x_2	x_3	x_4	x_1x_2	x_1x_3	x_1x_4	x_2x_3	x_2x_4	x_3x_4	$x_1x_2x_3$	$x_1x_2x_4$	$x_1x_3x_4$	$x_2x_3x_4$	$x_1x_2x_3x_4$	y
1	+	–	–	–	–	+	+	+	+	+	+	–	–	–	–	+	60.4
2	+	+	–	–	–	–	–	–	+	+	+	+	+	+	–	–	75.9
3	+	–	+	–	–	–	+	+	–	–	+	+	+	–	+	–	79.8
4	+	+	+	–	–	+	–	–	–	–	+	–	–	+	+	+	86.0
5	+	–	–	+	–	+	–	+	–	+	–	+	–	+	+	–	64.9
6	+	+	–	+	–	–	+	–	–	+	–	–	+	–	+	+	80.9
7	+	–	+	+	–	–	–	–	+	–	–	–	+	+	–	+	86.4
8	+	+	+	+	–	+	+	+	+	–	–	+	–	–	–	–	91.6
9	+	–	–	–	+	+	+	–	+	–	–	–	+	+	+	–	59.6
10	+	+	–	–	+	–	–	+	+	–	–	+	–	–	+	+	77.0
11	+	–	+	–	+	–	+	–	–	+	–	+	–	+	–	+	83.1
12	+	+	+	–	+	+	–	+	–	+	–	–	+	–	–	–	85.0
13	+	–	–	+	+	+	–	–	–	–	+	+	+	–	–	+	65.0
14	+	+	–	+	+	–	+	+	–	–	+	–	–	+	–	–	79.3
15	+	–	+	+	+	–	–	–	+	+	+	–	–	–	+	–	88.7
16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	91.1

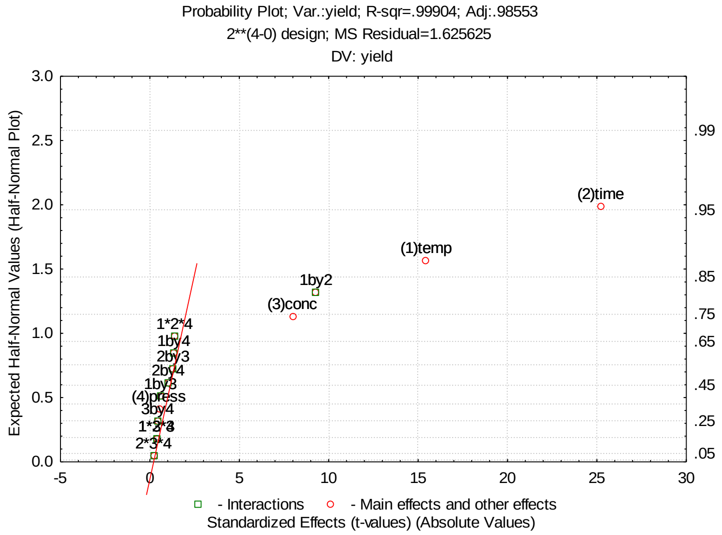


Effect Estimates; Var.:yield; R-sqr=.99904; Adj:.98553 (4fact)
2**(4-0) design; MS Residual=1.625625
DV: yield

Factor	Effect	Std.Err.	t(1)	p	-95.% Cnf.Limt	+95.% Cnf.Limt	Coeff.
Mean/Interc.	78.41875	0.318750	246.0196	0.002588	74.3686	82.46885	78.41875
(1)temp	9.86250	0.637500	15.4706	0.041093	1.7623	17.96271	4.93125
(2)time	16.08750	0.637500	25.2353	0.025214	7.9873	24.18771	8.04375
(3)conc	5.13750	0.637500	8.0588	0.078595	-2.9627	13.23771	2.56875
(4)press	0.36250	0.637500	0.5686	0.670847	-7.7377	8.46271	0.18125
1 by 2	-5.93750	0.637500	-9.3137	0.068092	-14.0377	2.16271	-2.96875
1 by 3	-0.38750	0.637500	-0.6078	0.652300	-8.4877	7.71271	-0.19375
1 by 4	-0.86250	0.637500	-1.3529	0.405214	-8.9627	7.23771	-0.43125
2 by 3	0.83750	0.637500	1.3137	0.414202	-7.2627	8.93771	0.41875
2 by 4	0.66250	0.637500	1.0392	0.487759	-7.4377	8.76271	0.33125
3 by 4	-0.28750	0.637500	-0.4510	0.730284	-8.3877	7.81271	-0.14375
1*2*3	0.26250	0.637500	0.4118	0.751332	-7.8377	8.36271	0.13125
1*2*4	-0.91250	0.637500	-1.4314	0.388215	-9.0127	7.18771	-0.45625
1*3*4	-0.26250	0.637500	-0.4118	0.751332	-8.3627	7.83771	-0.13125
2*3*4	0.16250	0.637500	0.2549	0.841108	-7.9377	8.26271	0.08125

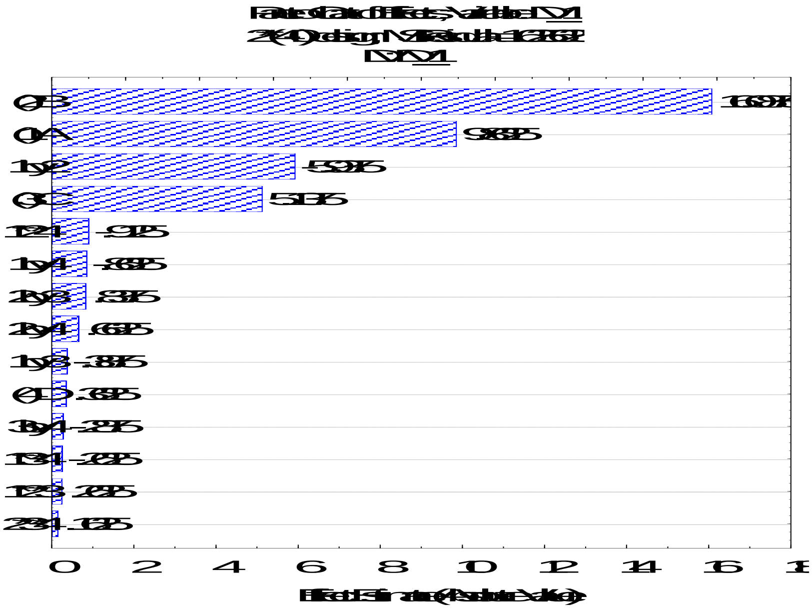
$b_0 = 78.42$; $b_1 = 4.93$; $b_2 = 8.04$; $b_3 = 2.57$; $b_4 = 0.18$; $b_{12} = –2.97$; $b_{13} = –0.19$; $b_{14} = –0.43$; $b_{23} = 0.42$; $b_{24} = –0.33$; $b_{34} = –0.14$; $b_{123} = 0.13$; $b_{124} = –0.46$; $b_{134} = –0.13$; $b_{234} = 0.08$; $b_{1234} = 0.32$

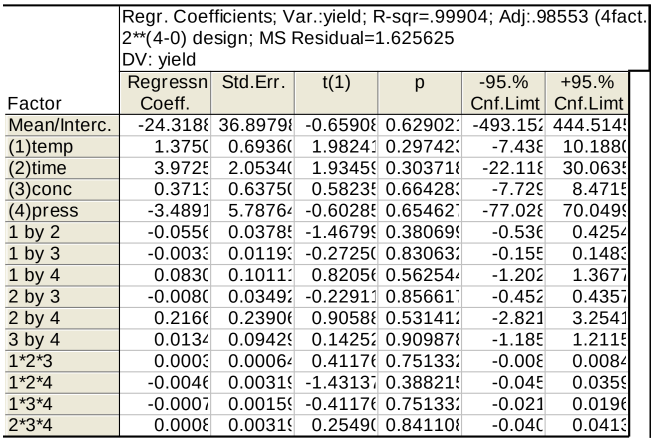



Probability Plot; Var.:yield; R-sqr=.99904; Adj:.98553
2**(4-0) design; MS Residual=1.625625
DV: yield

$$\hat{Y} = 78.42 + 4.93x_1 + 8.04x_2 + 2.53x_3 - 2.97x_1x_2$$

reduced model transfer function





Regr. Coefficients; Var.:yield; R-sqr=.99904; Adj:.98553 (4fact.
2**(4-0) design; MS Residual=1.625625
DV: yield

Factor	Regressn Coeff.	Std.Err.	t(1)	p	-95.% Cnf.Limt	+95.% Cnf.Limt
Mean/Interc.	-24.3188	36.89798	-0.65908	0.629021	-493.152	444.5145
(1)temp	1.3750	0.69360	1.98241	0.297423	-7.438	10.1880
(2)time	3.9725	2.05340	1.93459	0.303718	-22.118	30.0635
(3)conc	0.3713	0.63750	0.58235	0.664283	-7.729	8.4715
(4)press	-3.4891	5.78764	-0.60285	0.654627	-77.028	70.0499
1 by 2	-0.0556	0.03785	-1.46799	0.380699	-0.536	0.4254
1 by 3	-0.0033	0.01193	-0.27250	0.830632	-0.155	0.1483
1 by 4	0.0830	0.10111	0.82056	0.562544	-1.202	1.3677
2 by 3	-0.0080	0.03492	-0.22911	0.856617	-0.452	0.4357
2 by 4	0.2166	0.23906	0.90588	0.531412	-2.821	3.2541
3 by 4	0.0134	0.09429	0.14252	0.909878	-1.185	1.2115
1*2*3	0.0003	0.00064	0.41176	0.751332	-0.008	0.0084
1*2*4	-0.0046	0.00319	-1.43137	0.388215	-0.045	0.0359
1*3*4	-0.0007	0.00159	-0.41176	0.751332	-0.021	0.0196
2*3*4	0.0008	0.00319	0.25490	0.841108	-0.040	0.0413

Model parameters for the “non-transformed”, original variables (z_i).

- The estimated parameters depend on the factors included in the model because the design is not orthogonal on this scale.
- *p* is not informative

2^{p-r} fractional factorial designs

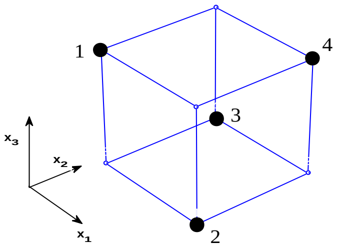
x_3

2^2				
i	x_0	x_1	x_2	x_1x_2
1	+	–	–	+
2	+	+	–	–
3	+	–	+	–
4	+	+	+	+

2^{3-1}				
i	x_0	x_1	x_2	x_3
1	+	–	–	+
2	+	+	–	–
3	+	–	+	–
4	+	+	+	+

2^2				
i	x_0	x_1	x_2	x_1x_2
1	+	–	–	+
2	+	+	–	–
3	+	–	+	–
4	+	+	+	+

2^{3-1}				
i	x_0	x_1	x_2	x_3
1	+	–	–	+
2	+	+	–	–
3	+	–	+	–
4	+	+	+	+



The model fitted

$$\hat{Y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3$$

i	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3
1	+	–	–	+	+	–	–
2	+	+	–	–	–	–	+
3	+	–	+	–	–	+	–
4	+	+	+	+	+	+	+

$b_3 \rightarrow \beta_3 + \beta_{12}$ since $x_3 = x_1 x_2$

Both sides are multiplied by x_3 $1 = x_1 x_2 x_3$

$x_1 = x_1^2 x_2 x_3 = x_2 x_3$ $b_1 \rightarrow \beta_1 + \beta_{23}$

$x_2 = x_1 x_3$ $b_2 \rightarrow \beta_2 + \beta_{13}$

(confounding pattern)

2^{4-1} $x_4 = x_1 x_2 x_3$ $1 = x_1 x_2 x_3 x_4$

Confounding/aliasing pattern

$x_1 = x_2 x_3 x_4$

$x_1 x_2 = x_3 x_4$

$x_2 = x_1 x_3 x_4$

$x_1 x_3 = x_2 x_4$

$x_3 = x_1 x_2 x_4$

$x_1 x_4 = x_2 x_3$

$x_4 = x_1 x_2 x_3$

Main effects are confounded with three-factor interactions, the two-factor interactions are confounded with each other

2^{5-1} $x_5 = x_1 x_2 x_3 x_4$ $1 = x_1 x_2 x_3 x_4 x_5$

$x_1 = x_2 x_3 x_4 x_5$ $x_1 x_2 = x_3 x_4 x_5$

2^{5-2} E.g. $x_4 = x_1 x_2$ $x_5 = x_1 x_2 x_3$

$$\hat{Y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + b_5 x_5$$

size of the design? number of parameters?

2^{5-3} 2^{6-3} 2^{7-3} 2^{7-4}

size of the design? number of parameters?

Example

G. E. P. Box, W. G. Hunter, J. S. Hunter: Statistics for Experimenters, J. Wiley, 1978; p. 424-429

	variable	-	+
1	water supply	town reservoir	well
2	raw material	on site	other
3	temperature	low	high
4	recycle	yes	no
5	caustic soda	fast	slow
6	filter cloth	new	old
7	holdup time	low	high

The first design:

				12	**13**	**23**	**123**	filtration time (min)
test	**1**	**2**	**3**	**4**	**5**	**6**	**7**	y
1	-	-	-	+	+	+	-	68.4
2	+	-	-	-	-	+	+	77.7
3	-	+	-	-	+	-	+	66.4
4	+	+	-	+	-	-	-	81.0
5	-	-	+	+	-	-	+	78.6
6	+	-	+	-	+	-	-	41.2
7	-	+	+	-	-	+	-	68.7
8	+	+	+	+	+	+	+	38.7

$x_4 = x_1 x_2$ $x_5 = x_1 x_3$ $x_6 = x_2 x_3$ $x_7 = x_1 x_2 x_3$

Analysis of the first design:

$l_1 = \boxed{-10.9} \rightarrow 1+24+35+67$

$l_2 = -2.8 \rightarrow 2+14+36+57$

$l_3 = \boxed{-16.6} \rightarrow 3+15+26+47$

$l_4 = 3.2 \rightarrow 4+12+37+56$

$l_5 = \boxed{-22.8} \rightarrow 5+13+27+46$

$l_6 = -3.4 \rightarrow 6+17+23+45$

$l_7 = 0.5 \rightarrow 7+16+25+34$

Second (fold-over) design:

				-12	**-13**	**-23**	**123**	filtration time (min)
test	**1**	**2**	**3**	**4**	**5**	**6**	**7**	y
1	+	+	+	-	-	-	+	66.7
2	-	+	+	+	+	-	-	65.0
3	+	-	+	+	-	+	-	86.4
4	-	-	+	-	+	+	+	61.9
5	+	+	-	-	+	+	-	47.8
6	-	+	-	+	-	+	+	59.0
7	+	-	-	+	+	-	+	42.6
8	-	-	-	-	-	-	-	67.6

Analysis of 16 runs:

$l_1 = -6.7 \rightarrow 1$
$l_2 = -3.9 \rightarrow 2$
$l_3 = -0.4 \rightarrow 3$
$l_4 = 2.8 \rightarrow 4$
$l_5 = -19.2 \rightarrow 5$
$l_6 = 0.1 \rightarrow 6$
$l_7 = -4.4 \rightarrow 7$
$l_{12} = 0.5 \rightarrow 12+37+56$
$l_{13} = -3.6 \rightarrow 13+27+46$
$l_{14} = 1.1 \rightarrow 14+36+57$
$l_{15} = -16.2 \rightarrow 15+26+47$
$l_{16} = 4.9 \rightarrow 16+25+34$
$l_{17} = -3.4 \rightarrow 17+23+45$
$l_{24} = -4.2 \rightarrow 24+35+67$

How much may the number of runs be reduced?

At least the main effects should appear in the model, minimally p+1 runs for p factors

E.g. 8 runs for 7 factors (2^{7-4}) .

Having 8-15 factors the number of runs is at least 16

Realising the design

Randomisation

The identical circumstances may not be assured for the whole set of experiments (there is not enough raw material from one batch, the number of runs is too high for realising it in one day or in one equipment, there are changes within one day).
If the order of realisation is that of the generating the design, the first half of the runs belong to the first level of a factor, the second half of the runs belong to the second level. Thus the main effect is conatminated with the effect of time.

The order of experiments may be randomised, then the day, lot, etc. does not contaminate the effects (no confounding).

In some cases a lot of raw material is not sufficient for the whole set of experiments, or the runs may not be performed on the same day or on the same equipment. If randomisation is used, the confounding is avoided (the effect of the lot of raw material the effect of the day or equipment does not contaminate the effect of factors), but as the systematic changes are turned to random variation, the variance is increased and may cover important effects.

Blocking is the solution. Within a block homogeneous circumstances are assured (same day, same lot of raw material, same equipment).

Blocking

BLOCK

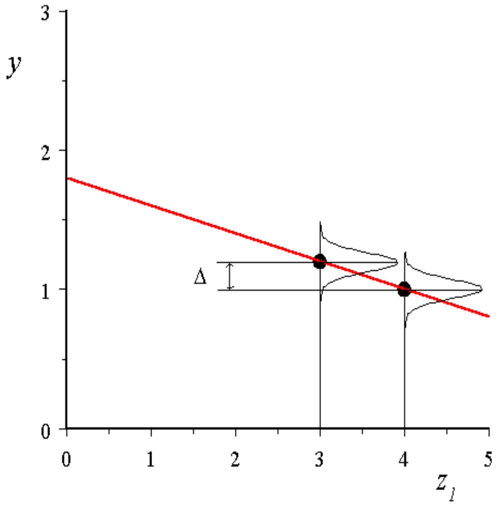
i	x_0	x_1	x_2	x_3	$x_1x_2x_3$
1	+	+	+	+	+
2	+	–	+	+	–
3	+	+	–	+	–
4	+	–	–	+	+
5	+	+	+	–	–
6	+	–	+	–	+
7	+	+	–	–	+
8	+	–	–	–	–

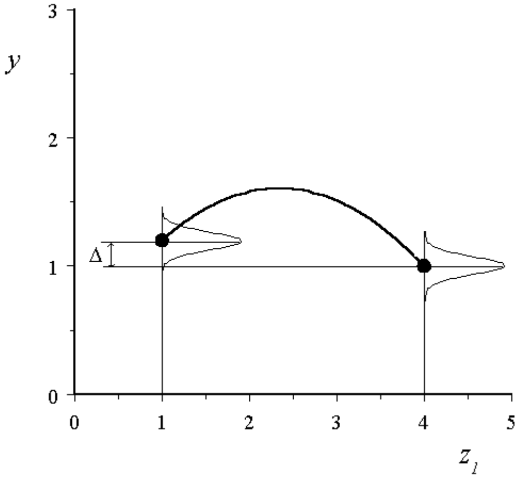
Choosing the width of interval of variation

Within the physical range

- the uncertainty of setting the factor levels should be negligible
- if the range is too narrow, the factor is qualified as insignificant
- if the range is too wide, the non-linearity may occur

The large variation covers the small effect




3
y
2
1
Δ
0
0
1
2
3
4
5
z_1