

## Lecture 10: $2^k$ Factorial Design

Montgomery: Chapter 6

## $2^k$ Factorial Design

- Involving  $k$  factors
- Each factor has two levels (often labeled + and -)
- Factor screening experiment (preliminary study)
- Identify important factors and their interactions
- Interaction (of any order) has **ONE** degree of freedom
- Factors need not be on numeric scale
- Ordinary regression model can be employed

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \epsilon$$

Where  $\beta_1$ ,  $\beta_2$  and  $\beta_{12}$  are related to main effects, interaction effects defined later.

## $2^2$ Factorial Design

Example:

factor		replicate				
A	B	treatment	1	2	3	mean
—	—	(1)	28	25	27	80/3
+	—	a	36	32	32	100/3
—	+	b	18	19	23	60/3
+	+	ab	31	30	29	90/3

- Let  $\bar{y}(A_+)$ ,  $\bar{y}(A_-)$ ,  $\bar{y}(B_+)$  and  $\bar{y}(B_-)$  be the level means of A and B.
- Let  $\bar{y}(A_-B_-)$ ,  $\bar{y}(A_+B_-)$ ,  $\bar{y}(A_-B_+)$  and  $\bar{y}(A_+B_+)$  be the treatment means

Define main effects of A (denoted again by A ) as follows:

$$\begin{aligned}
 A &= m.e.(A) = \bar{y}(A_+) - \bar{y}(A_-) \\
 &= \frac{1}{2}(\bar{y}(A_+B_+) + \bar{y}(A_+B_-)) - \frac{1}{2}(\bar{y}(A_-B_+) + \bar{y}(A_-B_-)) \\
 &= \frac{1}{2}(\bar{y}(A_+B_+) + \bar{y}(A_+B_-) - \bar{y}(A_-B_+) - \bar{y}(A_-B_-)) \\
 &= \frac{1}{2}(-\bar{y}(A_-B_-) + \bar{y}(A_+B_-) - \bar{y}(A_-B_+) + \bar{y}(A_+B_+)) \\
 &= 8.33
 \end{aligned}$$

- Let  $C_A = (-1, 1, -1, 1)$ , a contrast on treatment mean responses, then

$$m.e.(A) = \frac{1}{2} \hat{C}_A$$

- Notice that

$$A = m.e.(A) = (\bar{y}(A_+) - \bar{y}_{..}) - (\bar{y}(A_-) - \bar{y}_{..}) = \hat{\tau}_2 - \hat{\tau}_1$$

Main effect is defined in a different way than Chapter 5. But they are connected and equivalent.

- Similarly

$$B = m.e.(B) = \bar{y}(B_+) - \bar{y}(B_-)$$

$$= \frac{1}{2}(-\bar{y}(A_- B_-) - \bar{y}(A_+ B_-)) + \bar{y}(A_- B_+) + \bar{y}(A_+ B_+) = -5.00$$

Let  $C_B = (-1, -1, 1, 1)$ , a contrast on treatment mean responses, then  $B = m.e.(B) = \frac{1}{2} \hat{C}_B$

- Define interaction between A and B

$$AB = \text{Int}(AB) = \frac{1}{2}(m.e.(A | B_+) - m.e.(A | B_-))$$

$$= \frac{1}{2}(\bar{y}(A_+ | B_+) - \bar{y}(A_- | B_+)) - \frac{1}{2}(\bar{y}(A_+ | B_-) - \bar{y}(A_- | B_-))$$

$$= \frac{1}{2}(\bar{y}(A_- B_-) - \bar{y}(A_+ B_-) - \bar{y}(A_- B_+) + \bar{y}(A_+ B_+)) = 1.67$$

Let  $C_{AB} = (1, -1, -1, 1)$ , a contrast on treatment means, then

$$AB = \text{Int}(AB) = \frac{1}{2} \hat{C}_{AB}$$

## Effects and Contrasts

factor		effect (contrast)					
A	B	total	mean	I	A	B	AB
—	—	80	80/3	1	-1	-1	1
+	—	100	100/3	1	1	-1	-1
—	+	60	60/3	1	-1	1	-1
+	+	90	90/3	1	1	1	1

- There is a one-to-one correspondence between effects and contrasts, and contrasts can be directly used to estimate the effects.
- For a effect corresponding to contrast  $c = (c_1, c_2, \dots)$  in  $2^2$  design

$$\text{effect} = \frac{1}{2} \sum_i c_i \bar{y}_i$$

where  $i$  is an index for treatments and the summation is over all treatments.

## Sum of Squares due to Effect

- Because effects are defined using contrasts, their sum of squares can also be calculated through contrasts.
- Recall for contrast  $c = (c_1, c_2, \dots)$ , its sum of squares is

$$SS_{\text{Contrast}} = \frac{(\sum c_i \bar{y}_i)^2}{\sum c_i^2 / n}$$

So

$$SS_A = \frac{(-\bar{y}(A_-B_-) + \bar{y}(A_+B_-) - \bar{y}(A_-B_+) + \bar{y}(A_+B_+))^2}{4/n} = 208.33$$

$$SS_B = \frac{(-\bar{y}(A_-B_-) - \bar{y}(A_+B_-) + \bar{y}(A_-B_+) + \bar{y}(A_+B_+))^2}{4/n} = 75.00$$

$$SS_{AB} = \frac{(\bar{y}(A_-B_-) - \bar{y}(A_+B_-) - \bar{y}(A_-B_+) + \bar{y}(A_+B_+))^2}{4/n} = 8.33$$

## Sum of Squares and ANOVA

- Total sum of squares:  $SS_T = \sum_{i,j,k} y_{ijk}^2 - \frac{y_{\dots}^2}{N}$
- Error sum of squares:  $SS_E = SS_T - SS_A - SS_B - SS_{AB}$
- ANOVA Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	$F_0$
$A$	$SS_A$	1	$MS_A$	
$B$	$SS_B$	1	$MS_B$	
$AB$	$SS_{AB}$	1	$MS_{AB}$	
Error	$SS_E$	$N - 3$	$MS_E$	
Total	$SS_T$	$N - 1$		



## SAS file and output

```
option noncenter
data one;
input A B resp;
datalines;
-1 -1 28
-1 -1 25
-1 -1 27
 1 -1 36
 1 -1 32
 1 -1 32
-1  1 18
-1  1 19
-1  1 23
 1  1 31
 1  1 30
 1  1 29
;
proc glm;
class A B;
```

```
model resp=A|B;
run;
```

```
-----
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	291.6666667	97.2222222	24.82	0.0002
Error	8	31.3333333	3.9166667		
Cor Total	11	323.0000000			
A	1	208.3333333	208.3333333	53.19	<.0001
B	1	75.0000000	75.0000000	19.15	0.0024
A*B	1	8.3333333	8.3333333	2.13	0.1828

## Analyzing $2^2$ Experiment Using Regression Model

Because every effect in  $2^2$  design, or its sum of squares, has one degree of freedom, it can be equivalently represented by a numerical variable, and regression analysis can be directly used to analyze the data. The original factors are not necessarily continuous.

Code the levels of factor A and B as follows

A	$x_1$	B	$x_2$
-	-1	-	-1
+	1	+	1

Fit regression model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \epsilon$$

The fitted model should be

$$y = \bar{y}_{..} + \frac{A}{2} x_1 + \frac{B}{2} x_2 + \frac{AB}{2} x_1 x_2$$

i.e. the estimated coefficients are half of the effects, respectively.

## SAS Code and Output

```
option noncenter
data one;
input x1 x2 resp;
x1x2=x1*x2;
datalines;
-1 -1 28
-1 -1 25
-1 -1 27
.....
 1  1 31
 1  1 30
 1  1 29
;
proc reg;
model resp=x1 x2 x1x2;
run
```

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	291.66667	97.22222	24.82	0.0002
Error	8	31.33333	3.91667		
Corrected Total	11	323.00000			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	27.50000	0.57130	48.14	<.0001
x1	1	4.16667	0.57130	7.29	<.0001
x2	1	-2.50000	0.57130	-4.38	0.0024
x1x2	1	0.83333	0.57130	1.46	0.1828

## $2^3$ Factorial Design

Bottling Experiment:

factor			response			
A	B	C	treatment	1	2	total
-	-	-	(1)	-3	-1	-4
+	-	-	a	0	1	1
-	+	-	b	-1	0	-1
+	+	-	ab	2	3	5
-	-	+	c	-1	0	-1
+	-	+	ac	2	1	3
-	+	+	bc	1	1	2
+	+	+	abc	6	5	11

## factorial effects and constrasts

Main effects:

$$A = m.e.(A) = \bar{y}(A_+) - \bar{y}(A_-)$$

$$= \frac{1}{4}(\bar{y}(- - -) + \bar{y}(+ - -) - \bar{y}(- + -) + \bar{y}(+ + -) - \bar{y}(- - +) \\ + \bar{y}(+ - +) - \bar{y}(- + +) + \bar{y}(+ + +))$$

$$= 3.00$$

The contrast is (-1,1,-1,1,-1,1,-1,1)

$$B : (-1, -1, 1, 1, -1, -1, 1, 1), B = 2.25$$

$$C : (-1, -1, -1, -1, 1, 1, 1, 1), C = 1.75$$

2-factor interactions:

$$AB: A \times B \text{ componentwise, } AB=.75$$

$$AC: A \times C \text{ componentwise, } AC=.25$$

$$BC: B \times C \text{ componentwise, } BC=.50$$

3-factor interaction:

$$\begin{aligned}
 ABC = int(ABC) &= \frac{1}{2}(int(AB | C+) - int(AB | C-)) \\
 &= \frac{1}{4}(-\bar{y}(- - -) + \bar{y}(+ - -) + \bar{y}(- + -) - \bar{y}(+ + -) \\
 &\quad + \bar{y}(- - +) - \bar{y}(+ - +) - \bar{y}(- + +) + \bar{y}(+ + +)) \\
 &=.50
 \end{aligned}$$

The contrast is  $(-1,1,1,-1,1,-1,-1,1)= A \times B \times C$ .



### Contrasts for Calculating Effects in $2^3$ Design

			factorial effects								
A	B	C	treatment	<i>I</i>	<i>A</i>	<i>B</i>	<i>AB</i>	<i>C</i>	<i>AC</i>	<i>BC</i>	<i>ABC</i>
—	—	—	(1)	1	-1	-1	1	-1	1	1	-1
+	—	—	a	1	1	-1	-1	-1	-1	1	1
—	+	—	b	1	-1	1	-1	-1	1	-1	-1
+	+	—	ab	1	1	1	1	-1	-1	-1	1
—	—	+	c	1	-1	-1	1	1	-1	-1	1
+	—	+	ac	1	1	-1	-1	1	1	-1	-1
—	+	+	bc	1	-1	1	-1	1	-1	1	-1
+	+	+	abc	1	1	1	1	1	1	1	1

Estimates:

$$\text{grand mean: } \frac{\sum \bar{y}_i}{2^3}$$

$$\text{effect : } \frac{\sum c_i \bar{y}_i}{2^{3-1}}$$

Contrast Sum of Squares:

$$SS_{\text{effect}} = \frac{(\sum c_i \bar{y}_i)^2}{2^3/n} = 2n(\text{effect})^2$$

Variance of Estimate

$$\text{Var}(\text{effect}) = \frac{\sigma^2}{n2^{3-2}}$$

t-test for effects (confidence interval approach)

$$\text{effect} \pm t_{\alpha/2, 2^k(n-1)} \text{S.E.}(\text{effect})$$

## Regression Model

Code the levels of factor A and B as follows

A	$x_1$	B	$x_2$	C	$x_3$
-	-1	-	-1	-	-1
+	1	+	1	+	1

Fit regression model

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

The fitted model should be

$$y = \bar{y}_{..} + \frac{A}{2} x_1 + \frac{B}{2} x_2 + \frac{C}{2} x_3 + \frac{AB}{2} x_1 x_2 + \frac{AC}{2} x_1 x_3 + \frac{BC}{2} x_2 x_3 + \frac{ABC}{2} x_1 x_2 x_3$$

i.e.  $\hat{\beta} = \frac{\text{effect}}{2}$ , and

$$\text{Var}(\hat{\beta}) = \frac{\sigma^2}{n2^k} = \frac{\sigma^2}{n2^3}$$

## SAS Code: Bottling Experiment

```
data bottle;
input A B C devi;
datalines;
-1 -1 -1 -3
-1 -1 -1 -1
 1 -1 -1  0
 1 -1 -1  1
-1  1 -1 -1
-1  1 -1  0
 1  1 -1  2
 1  1 -1  3
-1 -1  1 -1
-1 -1  1  0
 1 -1  1  2
 1 -1  1  1
-1  1  1  1
-1  1  1  1
 1  1  1  6
 1  1  1  5
```

```
;
proc glm;
class A B C; model devi=A|B|C;
output out=botone r=res p=pred;
run;
proc univariate data=botone pctldef=4;
var res; qqplot res / normal (L=1 mu=est sigma=est);
histogram res / normal; run;
proc gplot; plot res*pred/frame; run;

data bottlenew;
set bottle;
x1=A; x2=B; x3=C; x1x2=x1*x2; x1x3=x1*x3; x2x3=x2*x3;
x1x2x3=x1*x2*x3; drop A B C;

proc reg data=bottlenew;
model devi=x1 x2 x3 x1x2 x1x3 x2x3 x1x2x3;
```

## SAS output for Bottling Experiment

ANOVA Model:

Dependent Variable: devi

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	73.00000000	10.42857143	16.69	0.0003
Error	8	5.00000000	0.62500000		
CorTotal	15	78.00000000			
A	1	36.00000000	36.00000000	57.60	<.0001
B	1	20.25000000	20.25000000	32.40	0.0005
A*B	1	2.25000000	2.25000000	3.60	0.0943
C	1	12.25000000	12.25000000	19.60	0.0022
A*C	1	0.25000000	0.25000000	0.40	0.5447
B*C	1	1.00000000	1.00000000	1.60	0.2415
A*B*C	1	1.00000000	1.00000000	1.60	0.2415

Regression Model:

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	1.00000	0.19764	5.06	0.0010
x1	1	1.50000	0.19764	7.59	<.0001
x2	1	1.12500	0.19764	5.69	0.0005
x3	1	0.87500	0.19764	4.43	0.0022
x1x2	1	0.37500	0.19764	1.90	0.0943
x1x3	1	0.12500	0.19764	0.63	0.5447
x2x3	1	0.25000	0.19764	1.26	0.2415
x1x2x3	1	0.25000	0.19764	1.26	0.2415

## General $2^k$ Design

- $k$  factors:  $A, B, \dots, K$  each with 2 levels (+, -)
- consists of all possible level combinations ( $2^k$  treatments) each with  $n$  replicates
- Classify factorial effects:

type of effect	label	the number of effects
main effects (of order 1)	$A, B, C, \dots, K$	$k$
2-factor interactions (of order 2)	$AB, AC, \dots, JK$	$\binom{k}{2}$
3-factor interactions (of order 3)	$ABC, ABD, \dots, IJK$	$\binom{k}{3}$
...	...	...
k-factor interaction (of order $k$ )	$ABC \dots K$	$\binom{k}{k}$



- In total, how many effects?
- Each effect (main or interaction) has 1 degree of freedom  
full model (i.e. model consisting of all the effects) has  $2^k - 1$  degrees of freedom.
- Error component has  $2^k(n - 1)$  degrees of freedom (why?).
- One-to-one correspondence between effects and contrasts:
  - For main effect: convert the level column of a factor using  $- \Rightarrow -1$  and  $+ \Rightarrow 1$
  - For interactions: multiply the contrasts of the main effects of the involved factors, componentwisely.

## General $2^k$ Design: Analysis

- Estimates:

$$\text{grand mean} : \frac{\sum \bar{y}_i}{2^k}$$

For effect with contrast  $C = (c_1, c_2, \dots, c_{2^k})$ , its estimate is

$$\text{effect} = \frac{\sum c_i \bar{y}_i}{2^{(k-1)}}$$

- Variance

$$\text{Var}(\text{effect}) = \frac{\sigma^2}{n2^{k-2}}$$

what is the standard error of the effect?

- t-test for  $H_0$ : effect=0. Using the confidence interval approach,

$$\text{effect} \pm t_{\alpha/2, 2^k(n-1)} \text{S.E.}(\text{effect})$$

**Using ANOVA model:**

- Sum of Squares due to an effect, using its contrast,

$$SS_{\text{effect}} = \frac{\sum c_i \bar{y}_{i.}^2}{2^k / n} = n 2^{k-2} (\text{effect})^2$$

- $SS_T$  and  $SS_E$  can be calculated as before and a ANOVA table including SS due to the effects and  $SS_E$  can be constructed and the effects can be tested by  $F$ -tests.

**Using regression:**

- Introducing variables  $x_1, \dots, x_k$  for main effects, their products are used for interactions, the following regression model can be fitted

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \dots + \beta_{12\dots k} x_1 x_2 \cdots x_k + \epsilon$$

The coefficients are estimated by half of effects they represent, that is,

$$\hat{\beta} = \frac{\text{effect}}{2}$$

**Unreplicated  $2^k$  Design**  
Filtration Rate Experiment

factor				filtration
<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	
-	-	-	-	45
+	-	-	-	71
-	+	-	-	48
+	+	-	-	65
-	-	+	-	68
+	-	+	-	60
-	+	+	-	80
+	+	+	-	65
-	-	-	+	43
+	-	-	+	100
-	+	-	+	45
+	+	-	+	104
-	-	+	+	75
+	-	+	+	86
-	+	+	+	70
+	+	+	+	96

## Unreplicated $2^k$ Design

- No degree of freedom left for error component if full model is fitted.
- Formulas used for estimates and contrast sum of squares are given in Slides 26-27 with  $n=1$
- No error sum of squares available, cannot estimate  $\sigma^2$  and test effects in both the ANOVA and Regression approaches.
- **Approach 1:** pooling high-order interactions
  - Often assume 3 or higher interactions do not occur
  - Pool estimates together for error
  - Warning: may pool significant interaction

## Unreplicated $2^k$ Design

- Approach 2: Using the normal probability plot (QQ plot) to identify significant effects.
  - Recall

$$\text{Var}(\text{effect}) = \frac{\sigma^2}{2^{(k-2)}}$$

If the effect is not significant ( $=0$ ), then the effect estimate follows

$$N\left(0, \frac{\sigma^2}{2^{(k-2)}}\right)$$

- Assume all effects not significant, their estimates can be considered as a random sample from  $N\left(0, \frac{\sigma^2}{2^{(k-2)}}\right)$
- QQ plot of the estimates is expected to be a linear line
- Deviation from a linear line indicates significant effects

## Using SAS to generate QQ plot for effects

```

goption colors=(none);

data filter;
  do D = -1 to 1 by 2;do C = -1 to 1 by 2;
  do B = -1 to 1 by 2;do A = -1 to 1 by 2;
  input y @@;  output;
  end; end; end; end;
datalines;
45 71 48 65 68 60 80 65 43 100 45 104 75 86 70 96
;

data inter;                                /* Define Interaction Terms */
  set filter;
  AB=A*B; AC=A*C; AD=A*D; BC=B*C; BD=B*D; CD=C*D; ABC=AB*C; ABD=AB*D;
  ACD=AC*D; BCD=BC*D; ABCD=ABC*D;

proc glm data=inter;                        /* GLM Proc to Obtain Effects */
  class A B C D AB AC AD BC BD CD ABC ABD ACD BCD ABCD;
  model y=A B C D AB AC AD BC BD CD ABC ABD ACD BCD ABCD;

```



```

estimate 'A' A 1 -1; estimate 'AC' AC 1 -1;
run;

proc reg outest=effects data=inter; /* REG Proc to Obtain Effects */
  model y=A B C D AB AC AD BC BD CD ABC ABD ACD BCD ABCD;

data effect2; set effects;
drop y intercept _RMSE_;
proc transpose data=effect2 out=effect3;
data effect4; set effect3; effect=col1*2;
proc sort data=effect4; by effect;
proc print data=effect4;

                                /*Generate the QQ plot */
proc rank data=effect4 out=effect5 normal=blom;
  var effect; ranks neff;
proc print data=effect5;
symbol1 v=circle;
proc gplot data=effect5;
  plot effect*neff=_NAME_;
run;

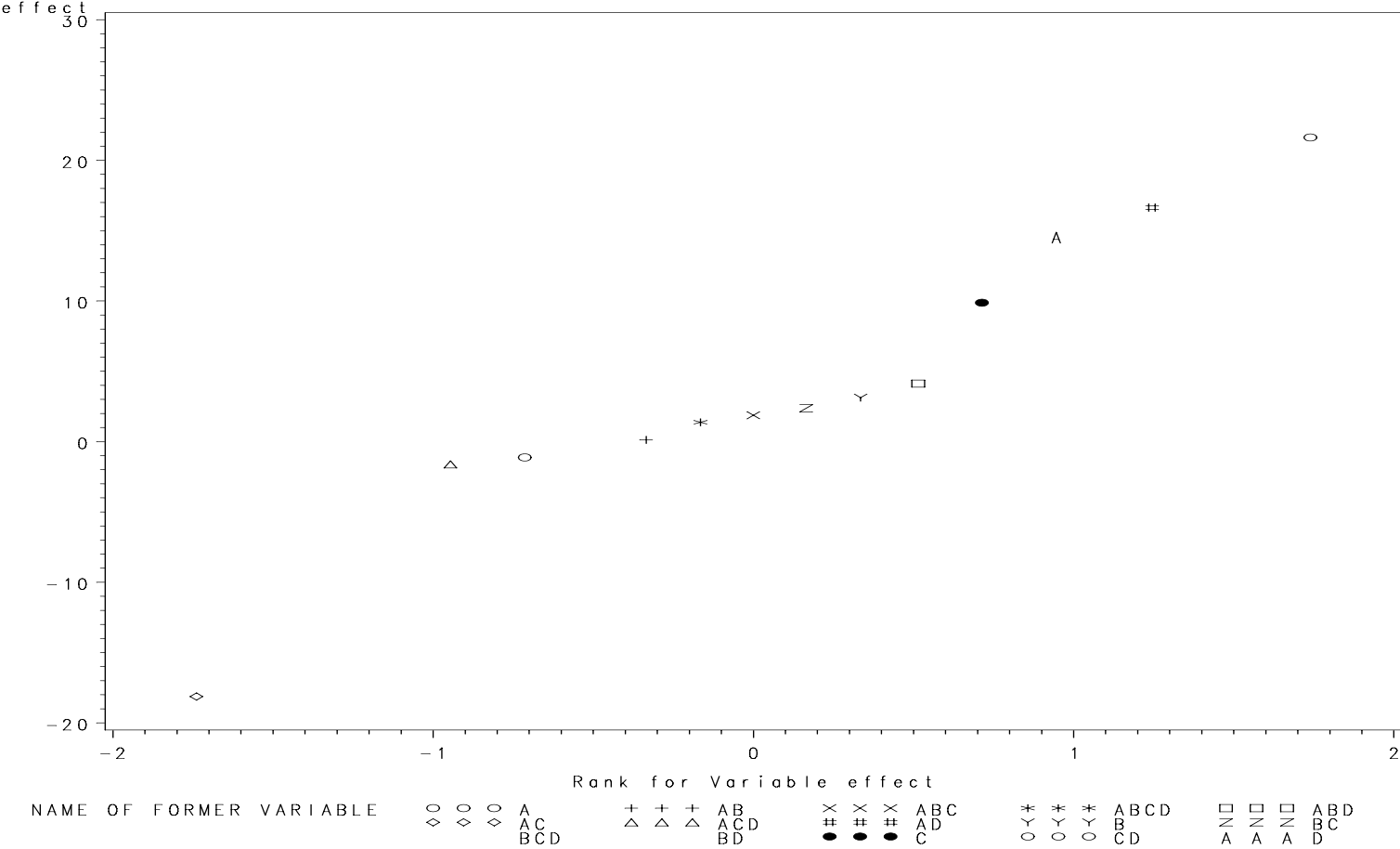
```

### Ranked Effects

Obs	_NAME_	COL1	effect	neff											
1	AC	-9.0625	-18.125	-1.73938											
2	BCD	-1.3125	-2.625	-1.24505											
3	ACD	-0.8125													
do D = -1 to 1 by 2; do C = -1 to 1 by 2;															
do B = -1 to 1 by 2; do A = -1 to 1 by 2;															
input y @@; output;															
end; end; end; end;															
datalines;															
45	71	48	65	68	60	80	65	43	100	45	104	75	86	70	96
		;5		-1.625		-0.94578									
4	CD	-0.5625	-1.125	-0.71370											
5	BD	-0.1875	-0.375	-0.51499											
6	AB	0.0625	0.125	-0.33489											
7	ABCD	0.6875	1.375	-0.16512											
8	ABC	0.9375	1.875	-0.00000											
9	BC	1.1875	2.375	0.16512											

10	B	1.5625	3.125	0.33489
11	ABD	2.0625	4.125	0.51499
12	C	4.9375	9.875	0.71370
13	D	7.3125	14.625	0.94578
14	AD	8.3125	16.625	1.24505
15	A	10.8125	21.625	1.73938

### QQ plot



## Filtration Experiment Analysis

Fit a linear line based on small effects, identify the effects which are potentially significant, then use ANOVA or regression fit a sub-model with those effects.

1. Potentially significant effects:  $A, AD, C, D, AC$ .
2. Use main effect plot and interaction plot
3. ANOVA model involving only  $A, C, D$  and their interactions (projecting the original unreplicated  $2^4$  experiment onto a replicated  $2^3$  experiment)
4. regression model only involving  $A, C, D, AC$  and  $AD$ .
5. Diagnostics using residuals.

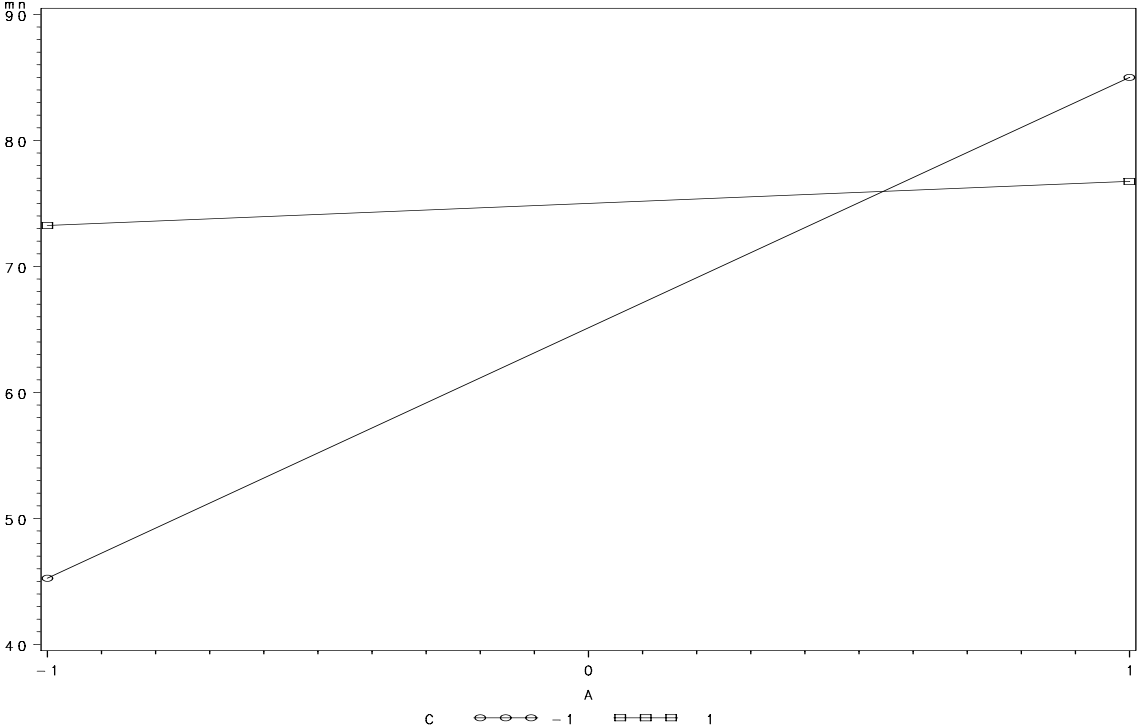
## Interaction Plots for $AC$ and $AD$

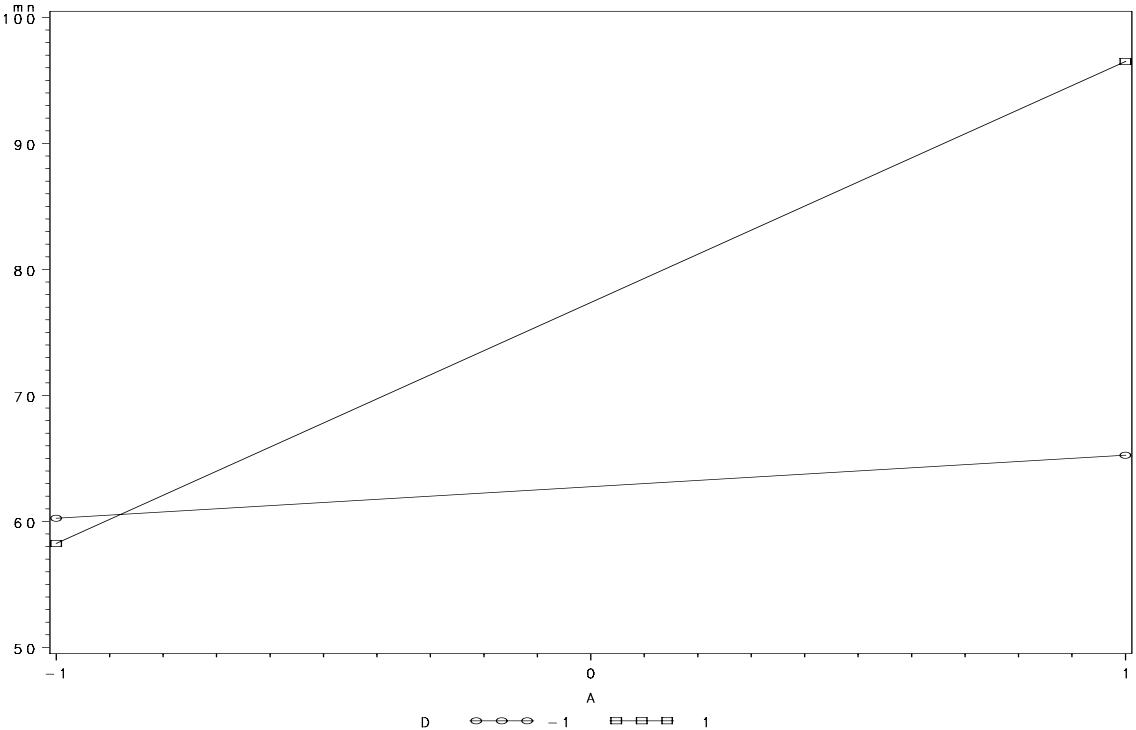
\* data step is the same.

```
proc sort; by A C;
proc means noprint;
var y; by A C;
output out=ymeanac mean=mn;

symbol1 v=circle i=join; symbol2 v=square i=join;
proc gplot data=ymeanac; plot mn*A=C;
run;
```

\* similar code for AD interaction plot







### ANOVA with *A*, *C* and *D* and their interactions

```
proc glm data=filter;
class A C D;
model y=A|C|D;
```

```
=====
```

Source	DF	Sum Squares	Mean Square	F Value	Pr > F
Model	7	5551.437500	793.062500	35.35	<.0001
Error	8	179.500000	22.437500		
Cor Total	15	5730.937500			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
A	1	1870.562500	1870.562500	83.37	<.0001
C	1	390.062500	390.062500	17.38	0.0031
A*C	1	1314.062500	1314.062500	58.57	<.0001
D	1	855.562500	855.562500	38.13	0.0003
A*D	1	1105.562500	1105.562500	49.27	0.0001
C*D	1	5.062500	5.062500	0.23	0.6475
A*C*D	1	10.562500	10.562500	0.47	0.5120

\*ANOVA confirms that A, C, D, AC and AD are significant effects

## Regression Model

\* the same data step

```
data inter; set filter; AC=A*C; AD=A*D;

proc reg data=inter; model y=A C D AC AD;
output out=outres r=res p=pred;

proc gplot data=outres; plot res*pred; run;
```

=====

Dependent Variable: y

### Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	5535.81250	1107.16250	56.74	<.0001
Error	10	195.12500	19.51250		
Corrected Total	15	5730.93750			
Root MSE		4.41730	R-Square	0.9660	

Dependent Mean            70.06250      Adj R-Sq            0.9489  
 Coeff Var                    6.30479

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	70.06250	1.10432	63.44	<.0001
A	1	10.81250	1.10432	9.79	<.0001
C	1	4.93750	1.10432	4.47	0.0012
D	1	7.31250	1.10432	6.62	<.0001
AC	1	-9.06250	1.10432	-8.21	<.0001
AD	1	8.31250	1.10432	7.53	<.0001

## Response Optimization / Best Setting Selection

Use  $x_1, x_3, x_4$  for  $A, C, D$ ; and  $x_1x_3, x_1x_4$  for  $AC, AD$  respectively. The regression model gives the following function for the response (filtration rate):

$$y = 70.06 + 10.81x_1 + 4.94x_3 + 7.31x_4 - 9.06x_1x_3 + 8.31x_1x_4$$

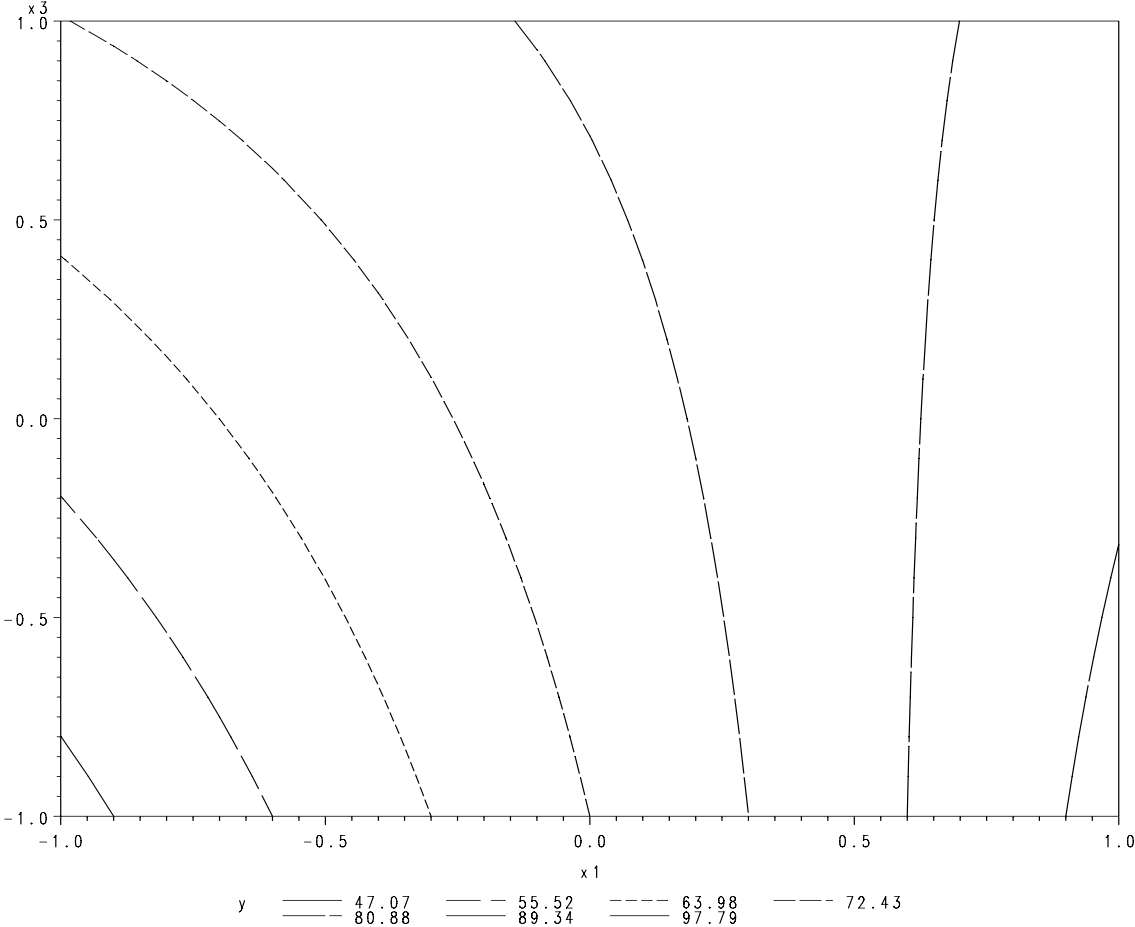
Want to maximize the response. Let  $D$  be set at high level ( $x_4 = 1$ )

$$y = 77.37 + 19.12x_1 + 4.94x_3 - 9.06x_1x_3$$

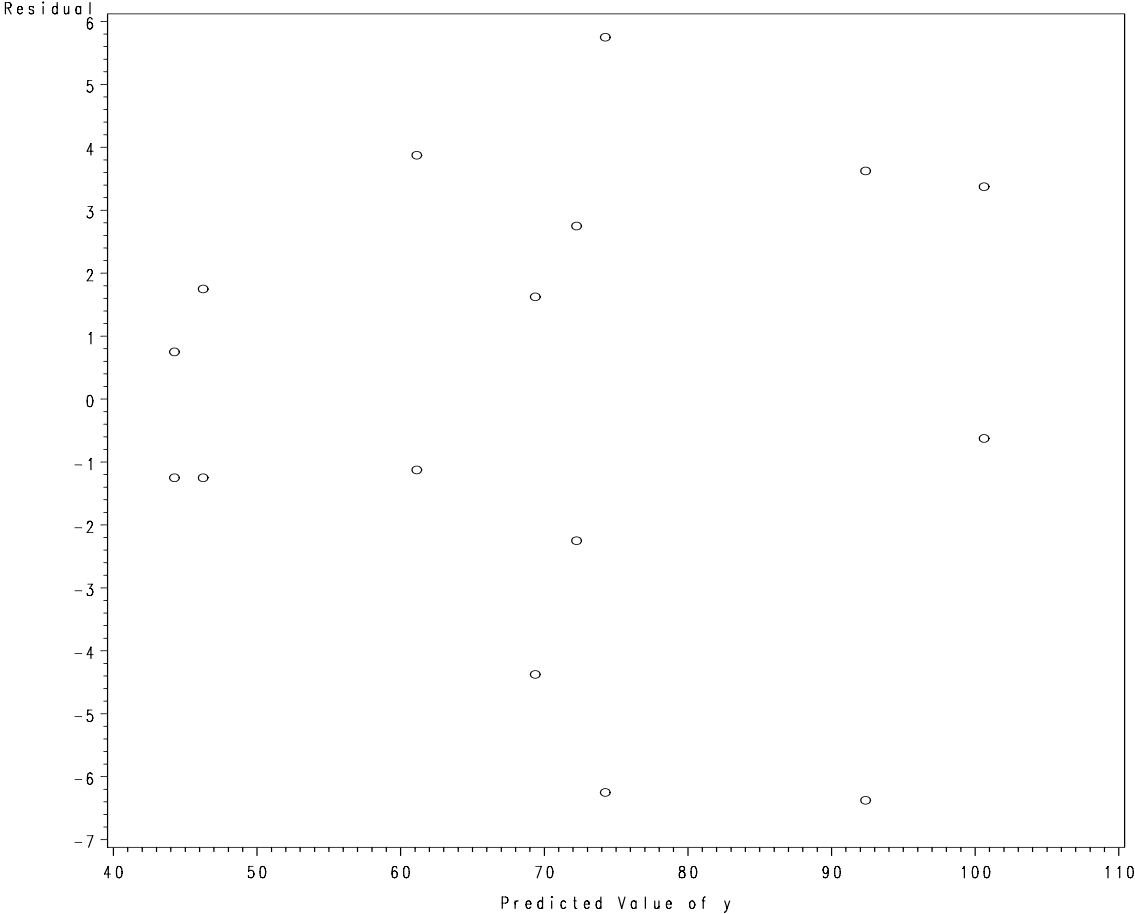
### Contour plot

```
goption colors=(none);
data one;
do x1 = -1 to 1 by .1;
  do x3 = -1 to 1 by .1;
    y=77.37+19.12*x1 +4.94*x3 -9.06*x1*x3 ; output;
  end; end;
proc gcontour data=one; plot x3*x1=y;
run; quit;
```

### Contour Plot for Response Given $D$



### Residual Plot



## Some Other Issues

- Half normal plot for  $(x_i), i = 1, \dots, n$ :
  - let  $\tilde{x}_i$  be the absolute values of  $x_i$
  - sort the  $(\tilde{x}_i)$ :  $\tilde{x}_{(1)} \leq \dots \leq \tilde{x}_{(n)}$
  - calculate  $u_i = \Phi^{-1}\left(\frac{n+i}{2n+1}\right), i = 1, \dots, n$
  - plot  $\tilde{x}_{(i)}$  against  $u_i$
  - look for a straight line

Half normal plot can also be used for identifying important factorial effects

- Other methods to identify significant factorial effects (Lenth method).  
Hamada&Balakrishnan (1998) analyzing unreplicated factorial experiments: a review with some new proposals, statistica sinica.
- Detect dispersion effects
- Experiment with duplicate measurements
  - for each treatment combination:  $n$  responses from duplicate

measurements

- calculate mean  $\bar{y}$  and standard deviation  $s$ .
- Use  $\bar{y}$  and treat the experiment as unreplicated in analysis