

Topic 14: 2^k factorial design (II)

Montgomery: chapter 6

Prof. Lingling An
University of Arizona

Outline

- General 2^k factorial design
- Unreplicated factorial design

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 $H_0: \text{effect}=0$. 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,

$$SS_{\text{effect}} = \frac{(\sum c_i \bar{y}_{i.})^2}{2^k/n} = n2^{k-2}(\text{effect})^2$$

- 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 5-6 with $n=1$
- No error sum of squares available, cannot estimate σ^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 - data

```
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
;
proc print data=filter;
run;
```

```
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;
run;
proc print data=inter;
run;
```

Obs	D	C	B	A	y	AB	AC	AD	BC	BD	CD	ABC	ABD	ACD	BCD	ABCD
1	-1	-1	-1	-1	45	1	1	1	1	1	1	-1	-1	-1	-1	1
2	-1	-1	-1	1	71	-1	-1	-1	1	1	1	1	1	1	-1	-1
3	-1	-1	1	-1	48	-1	1	1	-1	-1	1	1	1	-1	1	-1
4	-1	-1	1	1	65	1	-1	-1	-1	-1	1	-1	-1	1	1	1
5	-1	1	-1	-1	68	1	-1	1	-1	1	-1	1	-1	1	1	-1
6	-1	1	-1	1	60	-1	1	-1	-1	1	-1	-1	1	-1	1	1
7	-1	1	1	-1	80	-1	-1	1	1	-1	-1	-1	1	1	-1	1
8	-1	1	1	1	65	1	1	-1	1	-1	-1	1	-1	-1	-1	-1
9	1	-1	-1	-1	43	1	1	-1	1	-1	-1	-1	1	1	1	-1
10	1	-1	-1	1	100	-1	-1	1	1	-1	-1	1	-1	-1	1	1
11	1	-1	1	-1	45	-1	1	-1	-1	1	-1	1	-1	1	-1	1
12	1	-1	1	1	104	1	-1	1	-1	1	-1	-1	1	-1	-1	-1
13	1	1	-1	-1	75	1	-1	-1	-1	-1	1	1	1	-1	-1	1
14	1	1	-1	1	86	-1	1	1	-1	-1	1	-1	-1	1	-1	-1
15	1	1	1	-1	70	-1	-1	-1	1	1	1	-1	-1	-1	1	-1
16	1	1	1	1	96	1	1	1	1	1	1	1	1	1	1	1

```

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;
run;

```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	15	5730.937500	382.062500	.	.

Error	0	Source	DF	Type III SS	Mean Square	F Value	Pr > F
Corrected Total	15						
		A	1	1870.562500	1870.562500	.	.
		B	1	39.062500	39.062500	.	.
		C	1	390.062500	390.062500	.	.
		D	1	855.562500	855.562500	.	.
		AB	1	0.062500	0.062500	.	.
		AC	1	1314.062500	1314.062500	.	.
		AD	1	1105.562500	1105.562500	.	.
		BC	1	22.562500	22.562500	.	.
		BD	1	0.562500	0.562500	.	.
		CD	1	5.062500	5.062500	.	.
		ABC	1	14.062500	14.062500	.	.
		ABD	1	68.062500	68.062500	.	.
		ACD	1	10.562500	10.562500	.	.

Parameter	Estimate	Standard Error	t Value	Pr > t
A	21.6250000	.	.	.
AC	-18.1250000	.	.	.

Output (Proc reg)

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	70.06250	.	.	.
A	1	10.81250	.	.	.
B	1	1.56250	.	.	.
C	1	4.93750	.	.	.
D	1	7.31250	.	.	.
AB	1	0.06250	.	.	.
AC	1	-9.06250	.	.	.
AD	1	8.31250	.	.	.
BC	1	1.18750	.	.	.
BD	1	-0.18750	.	.	.
CD	1	-0.56250	.	.	.
ABC	1	0.93750	.	.	.
ABD	1	2.06250	.	.	.
ACD	1	-0.81250	.	.	.
BCD	1	-1.31250	.	.	.
ABCD	1	0.68750	.	.	.

```
data effect2; set effects;
  drop y intercept _RMSE_;
run;

proc transpose data=effect2 out=effect3;
run;

data effect4; set effect3; effect=col1*2;
run;

proc sort data=effect4; by effect;
run;

proc print data=effect4;
run;
```

Obs	_NAME_	COL1	effect
1	AC	-9.0625	-18.125
2	BCD	-1.3125	-2.625
3	ACD	-0.8125	-1.625
4	CD	-0.5625	-1.125
5	BD	-0.1875	-0.375
6	AB	0.0625	0.125
7	ABCD	0.6875	1.375
8	ABC	0.9375	1.875
9	BC	1.1875	2.375
10	B	1.5625	3.125
11	ABD	2.0625	4.125
12	C	4.9375	9.875
13	D	7.3125	14.625
14	AD	8.3125	16.625
15	A	10.8125	21.625


```
proc rank data=effect4 out=effect5  
normal=blom;  
  var effect; ranks neff;  
run;
```

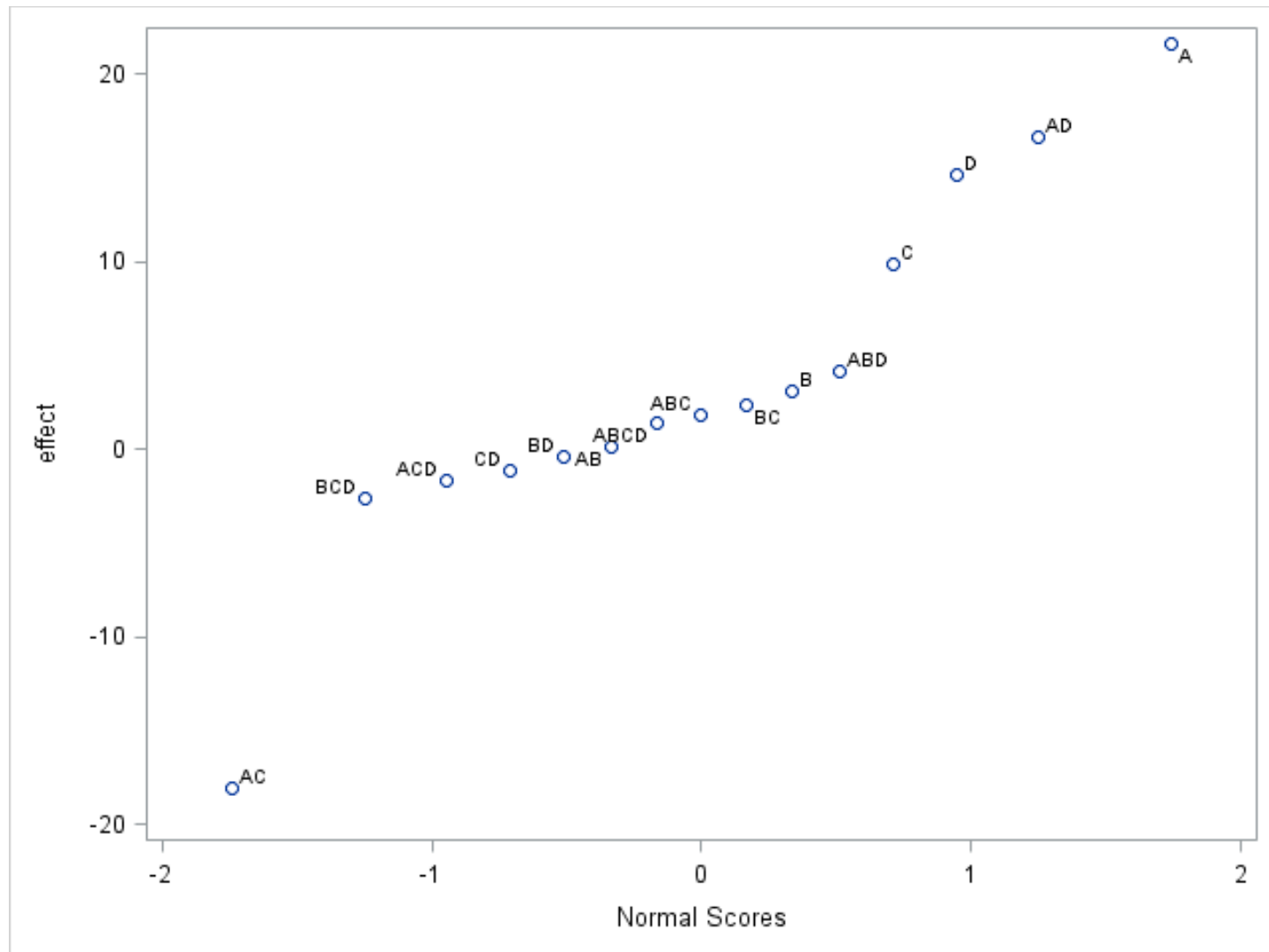
```
proc print data=effect5;  
run;
```

```
proc sgplot data=effect5;  
scatter x=neff y=effect/datalabel=_NAME_;  
xaxis label='Normal Scores';  
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	-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

Proc sgplot



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.

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

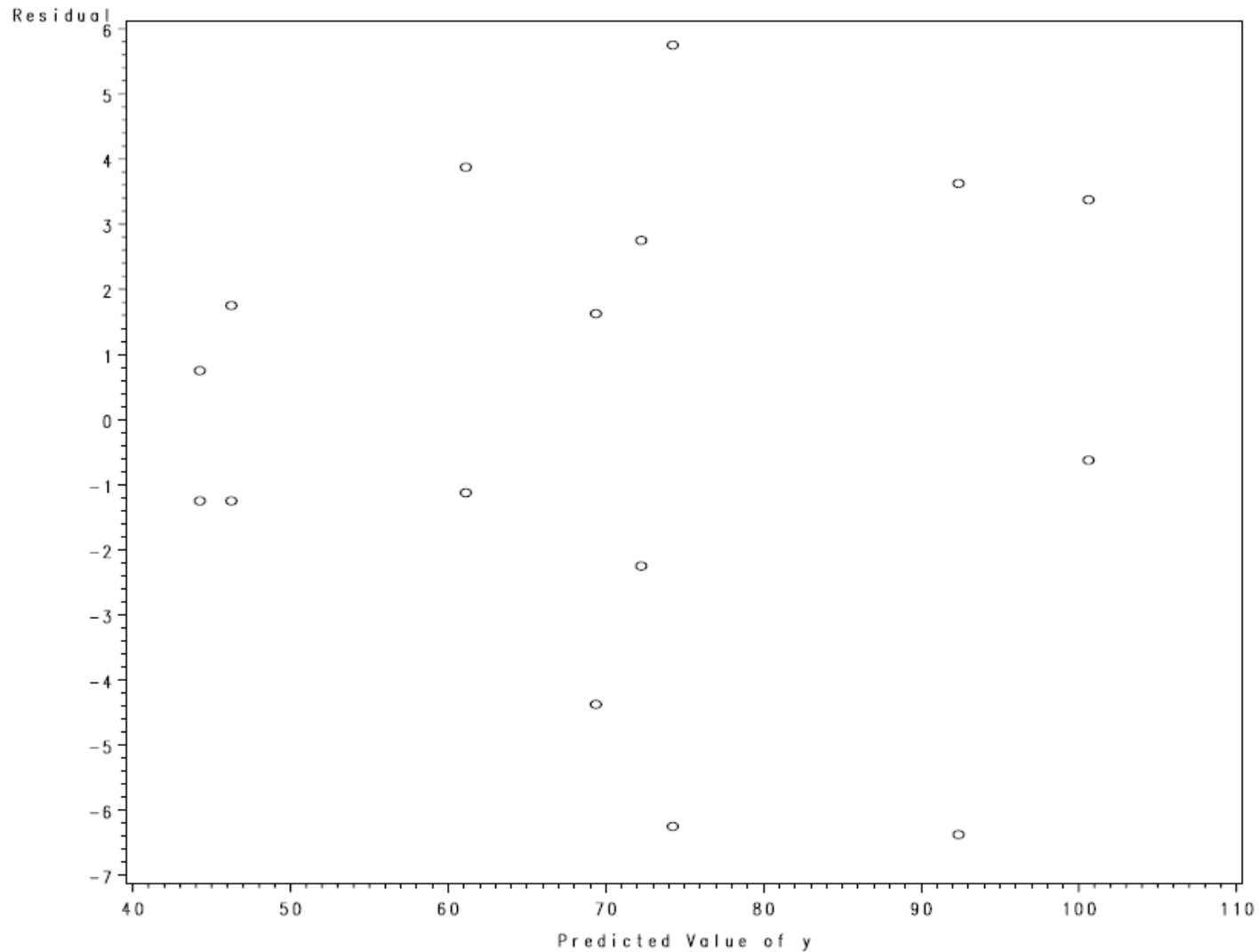
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	5535.81250	1107.16250	56.74	<.0001
Error	10	195.12500	19.51250		
Corrected Total	15	5730.93750			

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

Residual plot



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):

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

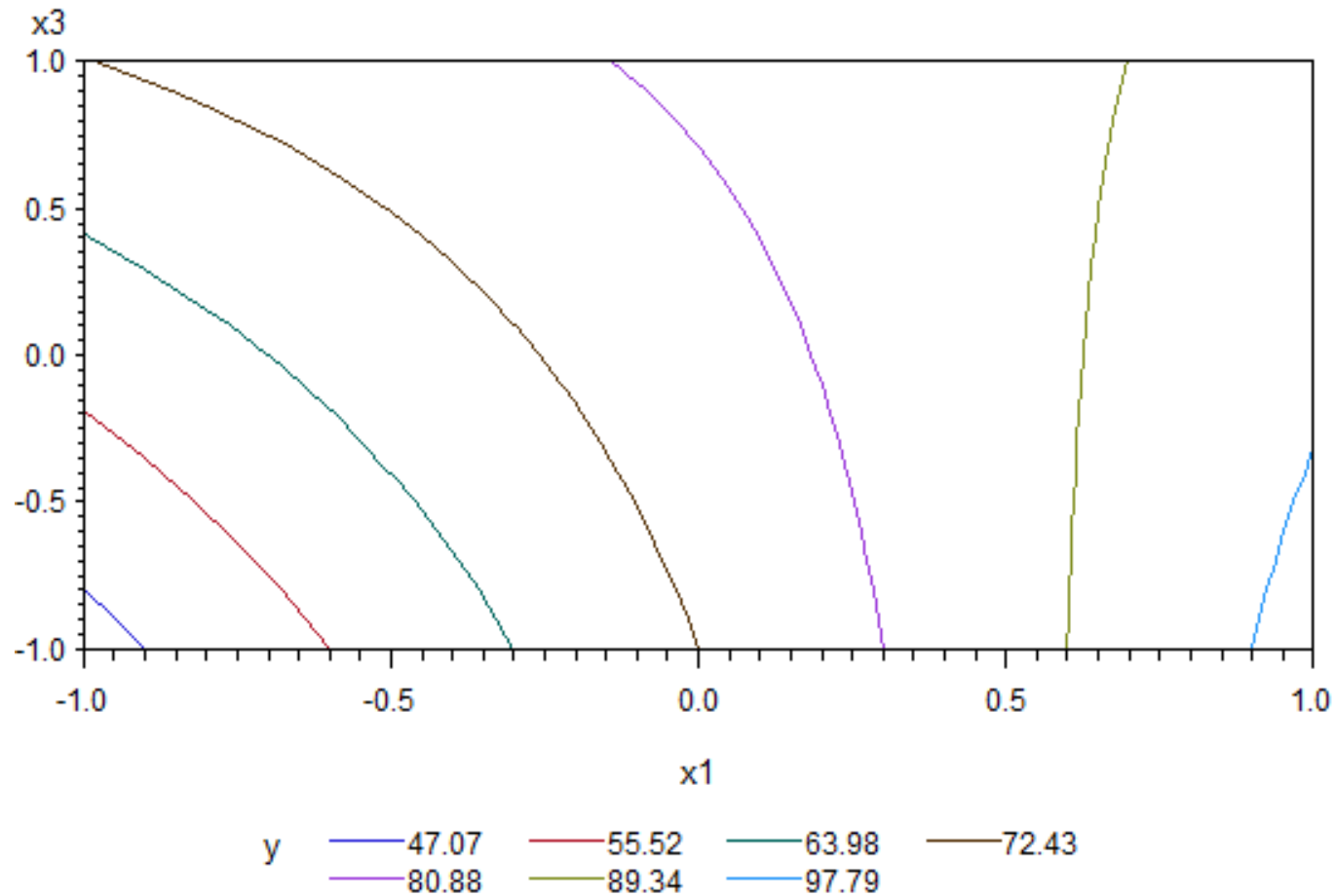
Want to maximize the response. Let D be set at high level ($x_4 = 1$)

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

Contour plot

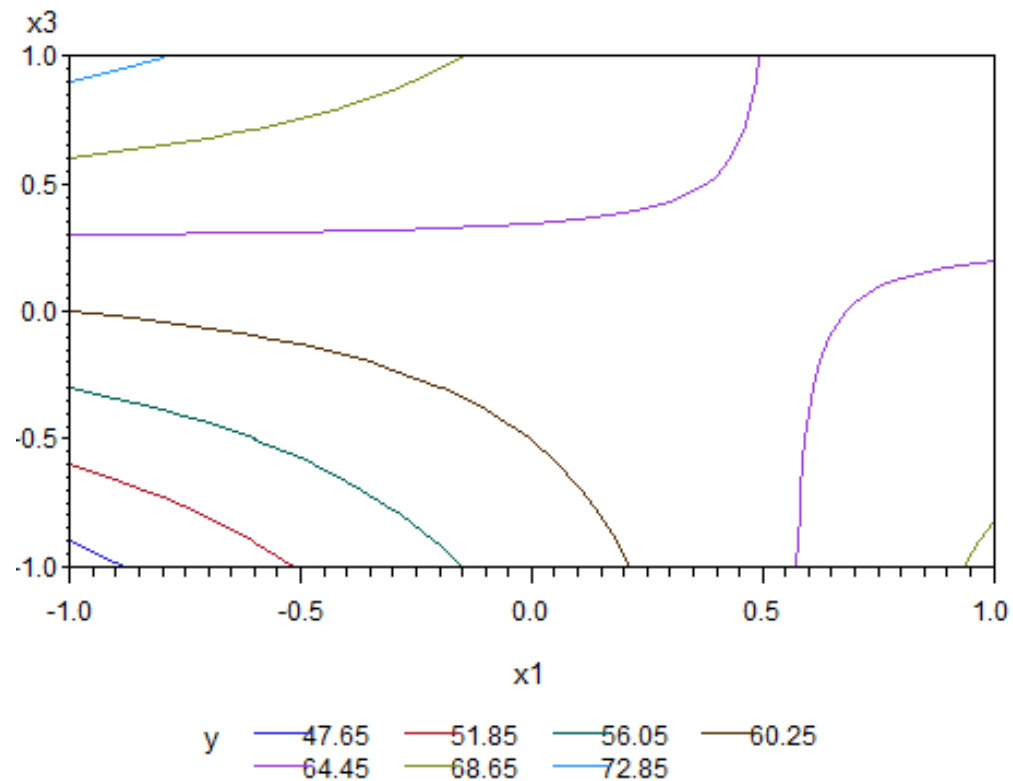
```
goption colors=(none);  
data one;  
do x1 = -1 to 1 by .1;  
  do x3 = -1 to 1 by .1;  
    y=77.37+19.12*x1 +4.94*x3 -9.06*x1*x3 ; output;  
  end; end;  
proc gcontour data=one; plot x3*x1=y;  
run; quit;
```

Contour Plot for Response Given $D (= +1)$



D= -1

```
data one;  
do x1 = -1 to 1 by .1;  
do x3 = -1 to 1 by .1;  
y=62.75+2.5*x1 +4.94*x3 -9.06*x1*x3 ; output;  
end; end;  
proc gcontour data=one; plot x3*x1=y;  
run;
```



Last slide

- Read Sections: 6.4-6.5

