Stat 591: Advanced Design of Experiments
Rutgers University
Fall 2001

Steve Buyske

Contents

LECTURE 1

Strategy of Experimentation I

"Experiments are an efficient way to learn about the world."
A. C. Atkinson

1. Introduction

- Text: Myers and Montgomery
- Computer: SAS
- Paper: oral presentation and short paper
- Take Home Final: Simulation
- Topics:
    1. Strategy of Experimentation
        (a) 2-level factorial and fractional factorial
        (b) Screening experiments
        (c) Response surface methodology
    2. Evolutionary Operation (EVOP)
    3. Optimal Design
    4. Computer-assisted design
    5. Mixture Experimentation
    6. Taguchi methods

2. Stategy of Experimentation

Typically, an experimenter is faced with a shopping list of factors, and must decide which ones are important. We would like to

1. Identify the primary factors.
2. Identify the region of interest for the primary factors.
3. Develop *useful* models in the region of the optimum settings of the factors.
4. Confirm the results.

3. Two Level Factorial Experiments

Perhaps the most common experimental technique, to the horror of statisticians, is to vary one factor at a time. For example, an experimenter might hold all factors fixed except for factor A, find the value for that factor that maximizes the response, and then repeat the process for factor B, and so on. We will see later why this is not a good plan.

Recall that a *factorial experiment* considers all possible treatments consisting of combinations of factors. For example if there are two factors with four and three levels, respectively, the experiment will have a total of $4 \times 3 = 12$ treatments.

In a two level factorial experiment, each factor is examined at *two levels*. These designs allow the examination of a relatively large number of factors with relatively few experimental runs. Why two levels? We need at least two to estimate the size of the effect, but using just two conserves resources.

They are restricted in terms of the size of the design space that can be explored.

They can be modified to reduce the number of the runs as well as increase the size of the design space.

They are excellent building blocks.

Three principles underlie our methods here. They are

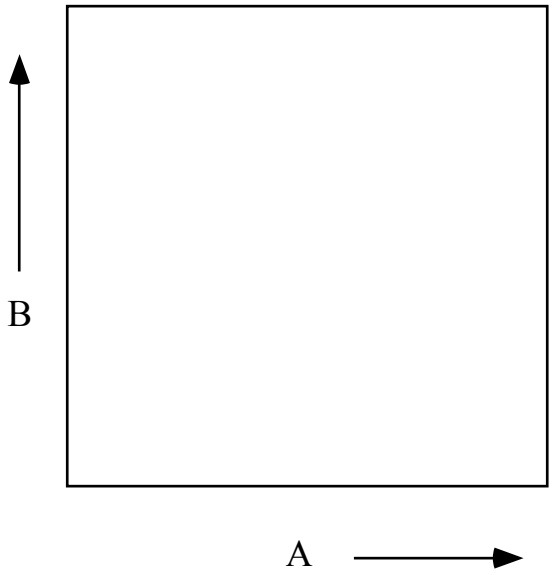
Hierarchical Ordering Principle: Lower order effects are more likely to be important than higher order effects.

Effect Sparsity, or Pareto, Principle: The number of important effects is generally small.

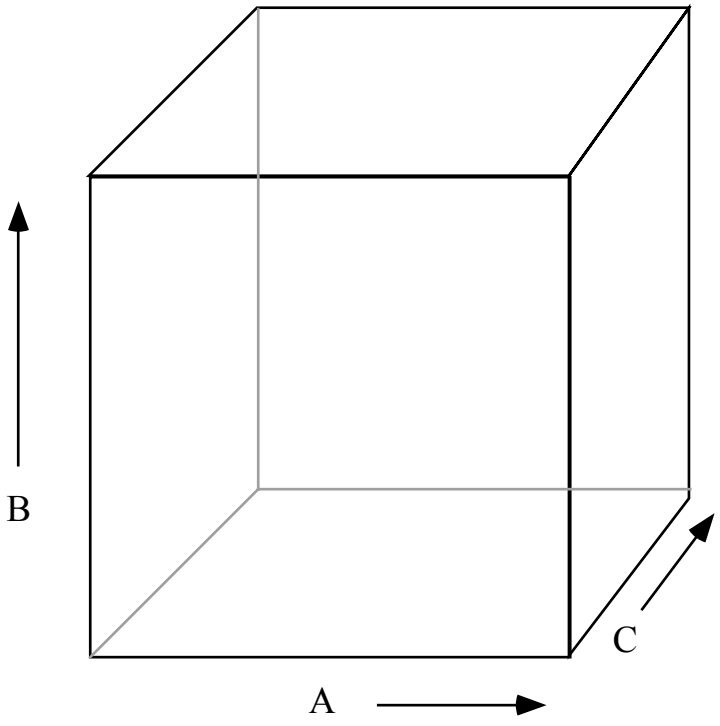
Effect Heredity Principle: For an interaction effect to be significant, (at least one of) its parent effects should be significant.

You can find a fuller description in C. F. J. Wu & M. Hamada, *Experiments: Planning, Analysis, and Parameter Design Optimization,* New York: John Wiley & Sons, Inc, 2000.

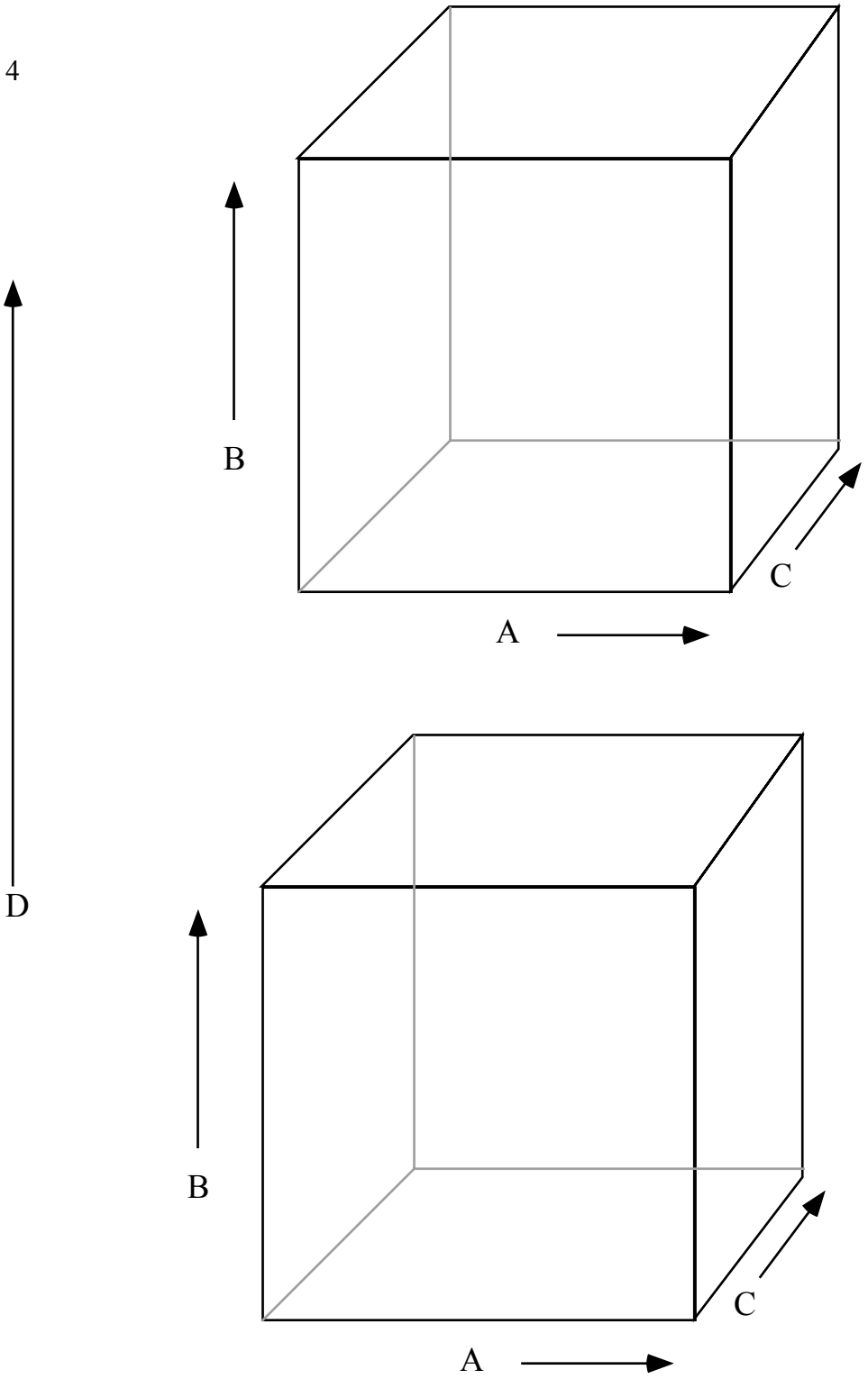
2^2



2^3



2^4



Let us examine a two-factor factorial so that we can set up some basic notation as well as introduce some fundamental ideas.

Notationally the designs are 2^K, where

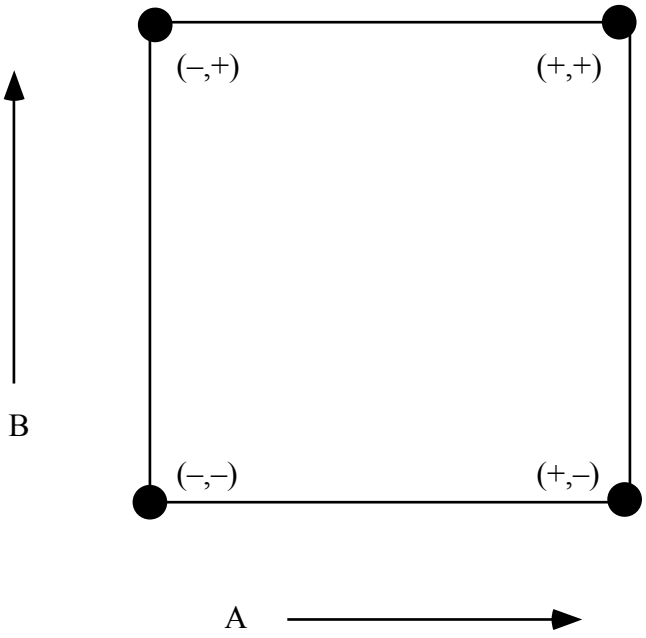
$$K = \text{number of factors},$$

and

2^K = number of treatment combination runs (usually just called "treatments").

For example if $K = 2$, the design requires $2^2 = 4$ runs.

Run	A	B
1	-	-
2	+	-
3	-	+
4	+	+



We may also denote these points using the *Yates notation*:

Run	Yates notation	A	B	data
1	(1)	-	-	2
2	a	+	-	4
3	b	-	+	6
4	ab	+	+	10

If we wanted to know the *effect of variable A* we could fix B, say $-$, and look at the difference between A_+ and A_-, i.e., $4 - 2 = 2$. We could then average this value with the difference when B is high, $10 - 6 = 4$. Therefore the *A effect* is

$$\frac{(4-2)+(10-6)}{2} = \frac{2+4}{2} = 3,$$

or, using the Yates notation,

$$\frac{(a-(1))+(ab-b)}{2}.$$

In general, we can determine how the effects are computed by constructing the following table:

	A	B	AB
(1)	-	-	+
a	+	-	-
b	-	+	-
ab	+	+	+

Therefore, except for the divisor,

$$\begin{aligned} \text{Mean} &= 1+a+b+ab \\ \text{A effect} &= -(1)+a-b+ab \\ \text{B effect} &= -(1)-a+b+ab \\ \text{AB effect} &= (1)-a-b+ab \end{aligned}$$

From the effects we can generate the SS for each factor (recall that the sum of squares for any contrast is the contrast squared divided by the number of observations times the sum of the squares of the contrast coefficients).

$$SS = \frac{(\text{numerator of effect})^2}{2^K}.$$

Therefore,

$$\begin{aligned} SS_A &= \frac{6^2}{4} = 9 \\ SS_B &= \frac{10^2}{4} = 25 \\ SS_{AB} &= \frac{2^2}{4} = 1 \\ SS_{TOT} &= 35. \end{aligned}$$

ANOVA

Source	df	SS	MS
Total	3	35	
A	1	9	9
B	1	25	25
AB	1	1	1

An alternative method to compute the effects and SS is to use the *Yates Algorithm*.

Order the data in the Yates standard notation.

Yates notation	data	col 1	col 2	effect= $\frac{\text{col K}}{2^{K-1}}$	SS= $\frac{(\text{col K})^2}{2^K}$
(1)	y_1	$y_1 + y_a$	$y_1 + y_a + y_b + y_{ab}$		
a	y_a	$y_b + y_{ab}$	$y_a - y_1 + y_{ab} - y_b$		
b	y_b	$y_a - y_1$	$y_b + y_{ab} - y_1 - y_a$		
ab	y_{ab}	$y_{ab} - y_b$	$y_{ab} - y_b - y_a + y_1$		

For the data

Yates notation	data	col 1	col 2	effect= $\frac{\text{col K}}{2^{K-1}}$	SS= $\frac{(\text{col K})^2}{2^K}$
(1)	2	$2 + 4 = 6$	$6 + 16 = 22 = \sum y$	$22/4 = 5.5 = \bar{\bar{y}}$	121=C.F.
a	4	$6 + 10 = 16$	$2 + 4 = 6$	3	9
b	6	$4 - 2 = 2$	$16 - 6 = 10$	5	25
ab	10	$10 - 6 = 4$	$4 - 2 = 2$	1	1

NB In the (1) row and the effect column, by dividing by 2^K instead of 2^{K-1}, we get the overall mean $\bar{\bar{y}}$.

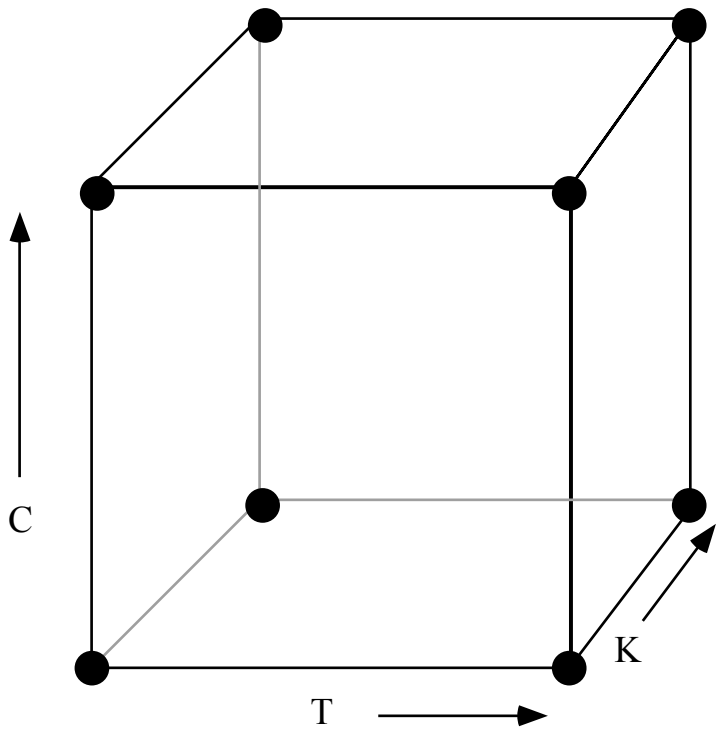
NB Because we have used -1 and +1 as the levels in our model, if you use a regression method to find the factor effects the coefficients you find will be one-half of that given here.

The Yates Algorithm is also described in Appendix 10A of Box, Hunter, and Hunter (on reserve).

Now let us consider a 2^3 *factorial experiment* given in Box, Hunter, and Hunter's *Statistics for Experimenters*.

We have two factors, Temperature and Concentration, that are quantitative, and Catalyst which is qualitative. The response is the chemical yield.

	Temperature T (– 160, + 180)	Concentration C (–20, + 40)	Catalyst K (– A, +B)	Yield data
(1)	–	–	–	60
t	+	–	–	72
c	–	+	–	54
tc	+	+	–	68
k	–	–	+	52
tk	+	–	+	83
ck	–	+	+	45
tck	+	+	+	80



Put data on corners of cube. Just by looking at the cube one can see a T effect, and C effect, and perhaps some sort of interaction.

To calculate the effect, we run the Yates algorithm.

		(1)	(2)	3	$(3)/2^{K-1}$	$(3)^2/2^K$
(1)	60	132	254	514	64.25	33024.5
t	72	122	260	92	23	1058
c	54	135	26	-20	-5	50
tc	68	125	66	6	1.5	4.5
k	52	12	-10	6	1.5	4.5
tk	83	14	-10	40	10	200
ck	45	31	2	0	0	0
tck	80	35	4	2	0.5	0.5
					Effect	SS

$$SS_{TOT} = 1317.5$$

Now, the problem remains how to determine which of the *factors is important*, "significant" in the scientific, not statistical, sense.

If we have duplicate observations within each cell then we can perform our usual test of significance.

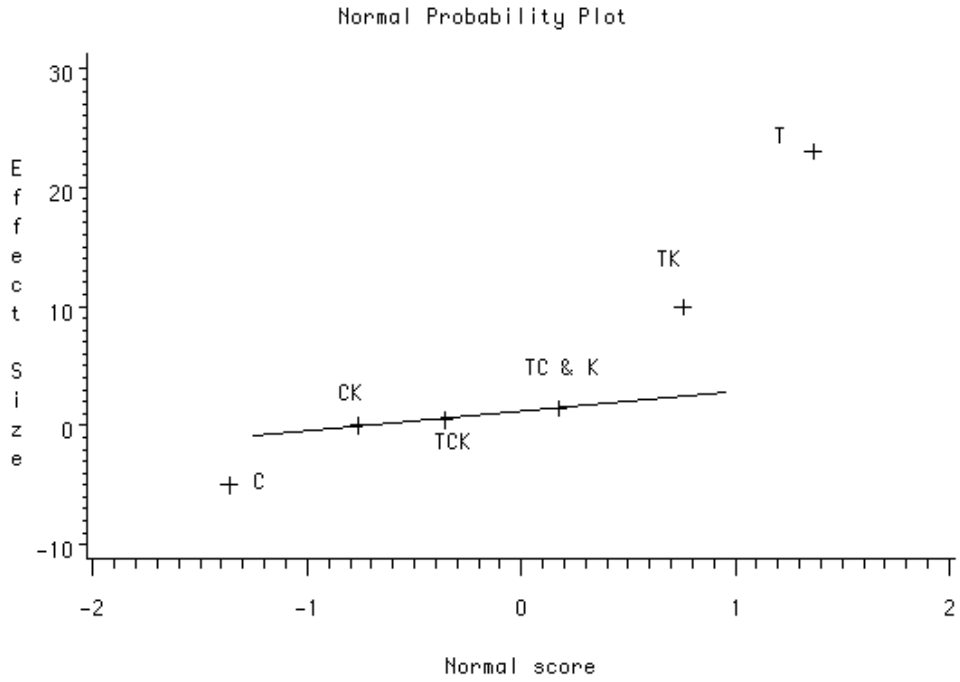
Many times, only one observation is collected within each cell. How can we handle the problem?

An alternative approach is to use normal probability plotting. To do this, first rank the effects. Next, calculate the cumulative percentiles. There are numerous

was to do this. The first column below is probably best, but the second and third are easier if you are doing it by hand or calculator. The conclusions that you draw will not be noticeably different.

Cum Percent				
$\frac{i-3/8}{\#\text{ of effects}+1/4} \times 100$	$\frac{i-1/2}{\#\text{ of effects}} \times 100$	$\frac{i}{\#\text{ of effects}} \times 100$	Effect	Source
8.6	7.1	12.5	-5	C
22.4	21.4	25.0	0	CK
36.2	35.7	37.5	0.5	TCK
56.9	50.0	56.25	1.5	TC
56.9	64.3	56.25	1.5	K
77.6	78.6	75.0	10	TK
91.4	92.9	87.5	23	T

NB The second column contains the values used in plotting on the next page.



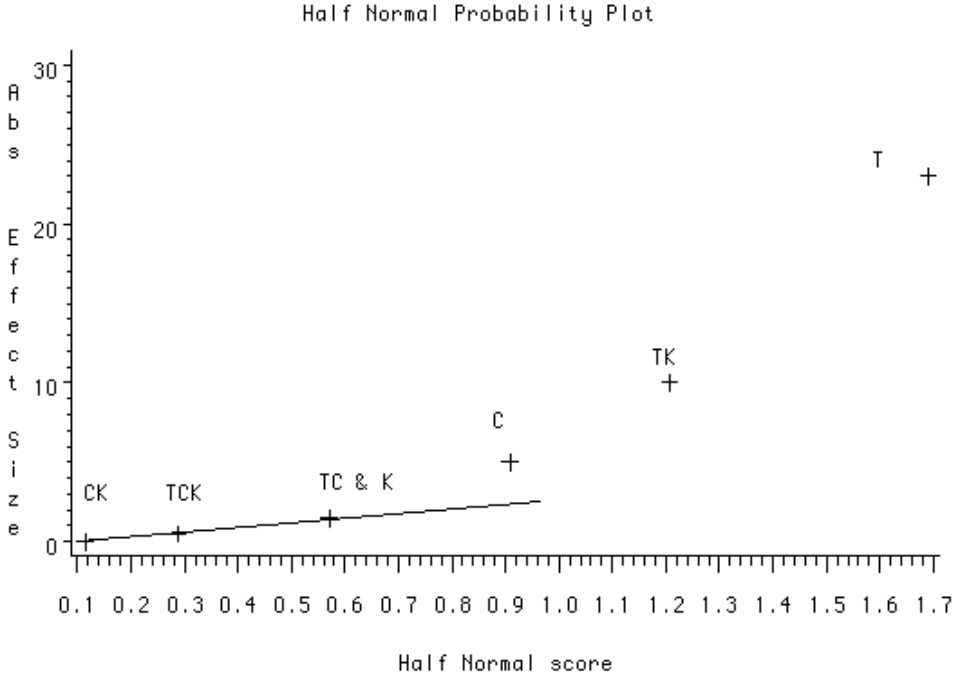
Discuss Plotting this info on normal probability paper.

The problem with this approach is that you need to look at both tails of the distribution as the sign of the effect is arbitrary as far as the importance of the effect is concerned. People tend to miss important effects this way.

An alternative approach is to use half-normal probability plotting. This paper can be purchased for varying values of p or it can be constructed from the usual normal probability paper. To do this, delete the scale for values of p less than 50%. For p greater than 50%, replace each value of P with P', where $P' = 2P - 100$. Now rank the absolute value of the effects and plot the data as was done earlier. People tend to see too many important effects this way.

Cum. Percent	|Effect|	Source
7.1	0	CK
21.4	0.5	TCK
35.7	1.5	TC
50.0	1.5	K
64.3	5	C
78.6	10	TK
92.9	23	T

See graph below.



```
/*
   SAS code for a
   demonstration of normal and half normal
   probability plots of effects from
   factorial experiments.
   Stat 591, Rutgers University
   Steve Buyske    Sept 99
*/

data effects;
    input effect @@;
    label effect='Effect Size';
    heffect=abs(effect);
    label heffect='Abs  Effect Size';
    cards;
    23 -5 1.5 1.5 10 0 .5;;

run;



/* calculate normal scores */
proc rank data=effects normal=blom out=normals;
   /* normal=blom means use (i-3/8)/(# of effects +1/4)*/
   var effect;
   ranks neffect;
run;

data normals;
set normals;
label neffect='Normal score';
run;

/* produce normal probability plot */
/* the lines commented out below give low-res plots */
/*
proc plot nolegend data=normals;
   title 'Normal Probability Plot';
   plot effect*neffect='*' ;
run;
*/

goptions ftext=none htext=1 cell;

proc gplot data=normals;
```

```
   title 'Normal Probability Plot';
   plot effect*neffect=effect;
run;

/* calculate half normal scores */
proc rank data=effects  out=hnranks;
   var heffect;
   ranks hneffect;
run;

data hnormals;
set hnranks nobs=n;
label hneffect='Half Normal score';
hneffect=probit(((hneffect-1/3)/(n+1/3))/2+.5);
run;

/* produce half normal probability plot */

/*
proc plot nolegend data=hnormals;
   title 'Normal Probability Plot';
   plot heffect*hneffect='*' ;
run;
*/


proc gplot data=hnormals;
    title 'Half Normal Probability Plot';
    plot heffect*hneffect=effect;
run;


goptions reset=all;
```

An alternative approach to the use of this probability plotting is to assume (for higher values of p) that all higher order interactions are not important (say 3-factor and higher). The SS for these terms may then be pooled together and be used as an estimate of the experimental error. This estimate may then be used as the divisor for the F test for the remaining factors. I personally do not recommend this approach, although it is popular in some quarters.

There are other formal methods of testing for important factors. A review of such methods can be found in M. Hamada & N. Balakrishnan, "Analyzing Unreplicated Factorial Experiments: A Review with Some New Proposals (with discussion)," *Statistica Sinica,* **8**, 1–41. The most popular is due to R. V. Lenth, "Quick and Easy Analysis of Unreplicated Factorials," *Technometrics* **31**, 469–473. First calculate the pseudo standard error,

$$PSE = 1.5\text{median}_{|\hat{\theta}_i|<2.5s_0}|\hat{\theta}_i|,$$

where

$$s_0 = 1.5\text{median}|\hat{\theta}_i|.$$

The PSE is a robust estimator the standard error of the non-active effects. Finally, one forms

$$t_{PSE,i} = \frac{\hat{\theta}_i}{PSE},$$

and compares to critical values in a table.

Extensions can be made of the *Yates algorithm*.

For example, the so called *reverse Yates algorithm* can be used to estimate the residuals. To see this consider the data we worked with earlier.

1. Take the last column prior to computing effects and SS, i.e., column 3 in this case.
2. For those factors felt to be n.s., replace the value shown with zero. In our example, replace TC, K, CK, TCK with zero. *Discuss this in terms of models*.
3. Reverse the order of the values from top to bottom.
4. Repeat the Yates algorithm.
5. Divide column 3 by 2^K giving $\hat{y}$.
6. Find the residuals $y - \hat{y}$.

Reverse Yates Algorithm

Effect	Reversed and modified col(3)	col 1	col 2	col 3	$\hat{y} = \frac{\text{col3}}{2^K=8}$	data (reverse also)	$y - \hat{y}$
TCK	0	0	40	626	78.25	80	1.75
CK	0	40	586	362	45.25	45	-0.25
TK	40	-20	-40	666	83.25	83	-0.25
K	0	606	402	402	50.25	52	1.75
TC	0	0	40	546	68.25	68	-0.25
C	-20	-40	626	442	55.25	54	-1.25
T	92	-20	-40	586	73.25	72	-1.25
(1)	514	422	442	482	60.25	60	-0.25

These residuals can be analyzed using techniques that you have learned earlier. In particular, you should look at a plot of residuals versus predicted, a plot of residuals versus observation number (or time), and a normal probability plot of the residuals.

The Yates algorithm can be used if more than one observation is observed for each treatment combination. In this case, cell totals are used in place of each observation. Effects and SS are then divided by n, the number of observations per cell. Everything is as usual except that the error SS will need to be computed by subtraction. (Note the Correction Factor is okay as well as all treatment effects.)

Homework. Read Chapter 1 of Myers and Montgomery, skim Chapter 2, and read Chapter 3. Additionally, do the handout.

LECTURE 2

Strategy of Experimentation II

Comments

- Computer Code.
- Last week's homework
- Interaction plots
- Helicopter project
-

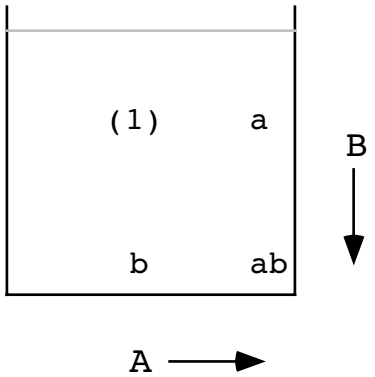
$$[4I \quad 2A \quad 2B \quad 2AB] = [\mu_{(1)} \quad \mu_A \quad \mu_B \quad \mu_{AB}] \begin{pmatrix} +1 & -1 & -1 & +1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & +1 & +1 \end{pmatrix}$$

1. Blocking

To this point we have assumed that the experimental units are sampled from a homogeneous population and that the experimental treatments are randomly assigned to the experimental units (*Completely randomized design*). Normally, when the units do not come from a homogeneous population, a *blocking* design is used. That type of design can also be used with 2^K factorial designs.

Consider the following problem.

We have a beaker of material which we need to sample. The factors of interest are the radius (center versus edge) and depth (top versus middle). The sampling procedure, however, may disturb the material. Furthermore, all four samples can't be removed simultaneously, as we have only two hands. Therefore, to obtain all four samples, the sampling must be done twice.



How should the samples be taken?
Suppose we took (1) and a first, b and ab second. Consider the effect table

	A	B	AB	block
(1)	-	-	+	1
a	+	-	-	1
b	-	+	-	2
ab	+	+	+	2

You can see that the blocking effect is *confounded* with the B effect. Can we avoid confounding the blocking effect with treatment effects? Consider the ANOVA table

Source	df
Total	3
A	1
B	1
AB	1
blocks	1

Since the total of degrees of freedom is 3, and we would like to partition it so that the total is 4, we clearly have a problem. Thus, some confounding is necessary. In this case, the best choice would be to confound the blocking effect with the AB effect so that the main effects are not confounded. Thus we should draw (1) and ab at the same time, and a and b at the same time. In doing so, the contrast $y_{(1)} - y_a - y_b + y_{ab}$ will estimate (twice) the AB effect plus the blocking effect (that's what we mean by confounding). Notationally we write Block $= AB$. Note that the A and B effects will be estimated independently of the blocking effect, because the contrasts are orthogonal.

Now consider a 2^3 factorial.

	A	B	AB	C	AC	BC	ABC
(1)	-	-	+	-	+	+	-
a	+	-	-	-	-	+	+
b	-	+	-	-	+	-	+
ab	+	+	+	-	-	-	-
c	-	-	+	+	-	-	+
ac	+	-	-	+	+	-	-
bc	-	+	-	+	-	+	-
abc	+	+	+	+	+	+	+

Think ANOVA

Source	df
Total	7
A	1
B	1
AB	1
C	1
AC	1
BC	1
ABC	1
Block	1
	?????

With two blocks, we might decide to confound blocks with ABC. Then treatments (1), ab, ac, and bc will be in one block, and a, b, c, and abc in the other.

Alternatively, we might need to use four blocks. To get started, we might confound two effects with blocks, say BC and ABC. In this case, (1) and bc would be in one block, a and abc in another, b and c in a third, and ab and ac in the fourth block. This design, however, has a serious weakness. Notice that this blocking pattern is also confounded with A. That is, we could have used A and BC to generate the block pairings. This should not surprise us as the degrees of freedom for blocks is 3.

The procedure to find confounding patterns is to pick out the two factors, BC and ABC and do binary arithmetic on the letters

$$(BC)(ABC) = AB^2C^2 = A.$$

Thus the third factor that is confounded is A.

An alternative procedure for blocking might be AB and BC giving $AB^2C = AC$ as the third factor confounded.

The resulting ANOVA table is then

Source	df
Total	7
A	1
B	1
C	1
ABC	1
Blocks, AB, AC, BC	3

[The SS may be computed from the Yates algorithm using the usual procedures, and with AB, AC, and BC being pooled together afterwords.] In general, with p blocks we will have $2^p - 1$ degrees of freedoms for blocks, and can estimate $2^k - 2^p$ unconfounded effects.

For this set up we have

Block	Treatment Run
1	(1), abc
2	a, bc
3	b, ac
4	ab,c

Incidentally, the block that contains treatment (1) is called the principal block. Notice that every other block could be obtained from the principal block by multiplying by the right letter(s) and using binary arithmetic.

With 2^K designs it is sometimes necessary to *replicate the study*.

E.g., $\boxed{2^3}$ $\boxed{2^3}$

Consider the ANOVA table.

Source	df
Total	15
A	1
B	1
AB	1
C	1
AC	1
BC	1
ABC	1
Block	1
A × Block	1
B × Block	1
AB × Block	1
C × Block	1
AC × Block	1
BC × Block	1
ABC × Block	1

Discuss direct tests versus combining these factors. Block is a random effect, so the proper F tests of the effects have $F(1, 1)$ distributions. The resulting tests have almost no power. The common practice is to pool all the Block×Factor interactions into the error term, thereby assuming that there is no Block×Factor interaction. The practical statistician might want to check for any unusual result first, however, by looking at residuals by block.

Now, if within each replication, *blocking is necessary*, what options do we have available?

1. We could use the *same blocking pattern* in each rep, thus completely confounding one factor with blocks.

E.g.,

Rep I		Rep II	
Blk 1	Blk 2	Blk 1	Blk 2
(1)	a	a	(1)
ab	b	b	ab
ac	c	c	ac
bc	abc	abc	bc

Here, Block $= ABC$. See below for the ANOVA table.

2. We could confound one factor in Rep I and a second factor in Rep II.
E.g.,

Rep I		Rep II	
Blk 1	Blk 2	Blk 1	Blk 2
(1)	a	(1)	b
ab	b	a	ab
ac	c	bc	c
bc	abc	abc	ac
Block $= ABC$		Block $= BC$	

These two different choices would give the following ANOVA tables.

Design 1		**Design 2**	
Source	df	Source	df
Total	15	Total	15
Rep	1	Rep	1
$ABC =$ Block(Rep) $+R \times ABC$	2		
A	1	A	1
B	1	B	1
AB	1	AB	1
C	1	C	1
AC	1	AC	1
BC	1	BC	$1'$
		ABC	$1'$
$A \times R$	1	$A \times R$	1
$B \times R$	1	$B \times R$	1
$AB \times R$	1	$AB \times R$	1
$C \times R$	1	$C \times R$	1
$AC \times R$	1	$AC \times R$	1
$BC \times R$	1	$BC \times R$	$1'$
		$ABC \times R$	$1'$
ABC totally confounded with blocks		ABC and BC partially confounded with blocks	

To analyze a partially confounded data, run the Yates algorithm separately on each replication. Then calculate each effect by averaging over the *unconfounded* estimates of that effect (i.e., omitting the estimate in any replication in which that estimate was confounded with the replication. For the SS, take the same approach with the square of the unconfounded effects.

References:

Box, Hunter, and Hunter. *Statistics for Experimenters*, Wiley, 1978.

Davies, O. L. *The Design and Analysis of Industrial Experiments*, Hafner, 1971.

Cochran and Cox, *Experimental Designs*, Wiley, 1957.

Daniel, C. *Applications of Statistics to Industrial Experimentation*, Wiley, 1976.

Daniel, C. Use of half-normal plots in interpreting factorial two-level experiments, *Technometrics* **1** 1959, pp 311–.

Daniel and Wood, *Fitting Equations to Data*, Wiley, 1980. Cox and Cox, *The Theory of the Design of Experiments,* Chapman & Hall/CRC, 2000.

2. Fractional Factorial Designs

With two-level factorial designs, when K becomes large, the number of experimental runs can become unmanageable. For example, if $K = 7$,

$$2^K = 2^7 = 128.$$

What do we get for all this work? The ability to estimate all main effects and interactions. For example, if $K = 7$

Interactions	number	
grand mean	1	$\binom{7}{0}$
main effects	7	$\binom{7}{1}$
2-way	21	$\binom{7}{2}$
3-way	35	$\binom{7}{3}$
4-way	35	$\binom{7}{4}$
5-way	21	$\binom{7}{5}$
6-way	7	$\binom{7}{6}$
7-way	1	$\binom{7}{7}$

The problem is that in most cases the higher order interactions are not likely to be real. Therefore we have expended a tremendous amount of effort and are not receiving full benefit.

	A	B	C	AB	AC	BC	ABC
(1)	-	-	-	+	+	+	-
a	+	-	-	-	-	+	+
b	-	+	-	-	+	-	+
ab	+	+	-	+	-	-	-
c	-	-	+	+	-	-	+
ac	+	-	+	-	+	-	-
bc	-	+	+	-	-	+	-
abc	+	+	+	+	+	+	+

If we confound Blocks=ABC, then
(1), ab, ac, bc go in block I
a, b, c, abc go in block II.

Suppose we only ran the data in block II

	A	B	C	AB	AC	BC	ABC
a	+	-	-	-	-	+	+
b	-	+	-	-	+	-	+
c	-	-	+	+	-	-	+
abc	+	+	+	+	+	+	+

The defining contrast or defining relation

$$I = ABC$$

gives

$$\begin{aligned} C &= ABCC = ABC^2 = AB(I) = AB \\ B &= AC \\ A &= BC \end{aligned}$$

Working our way forward to construct the design

$$\begin{aligned} C &= AB \\ I &= ABC \end{aligned}$$

		A	B	AB=C
(1)	c	-	-	+
a		+	-	-
b		-	+	-
ab	c	+	+	+

This gives the previous design, with A=BC, B=AC, and AB=C, all generated by the relation I=ABC.

Consider a 2^4 factorial

		y	Estimated Effect	
	(1)	71	72.25	
•	a	61	-8.00	$*$
•	b	90	24.00	$*$
	ab	82	1.00	
•	c	68	-2.25	$*$
	ac	61	0.75	
	bc	87	-1.25	
•	abc	80	-0.75	
•	d	61	-5.5	$*$
	ad	50	0.00	
	bd	89	4.50	$*$
•	abd	83	0.50	
	cd	59	-0.25	
•	acd	51	-0.25	
•	bcd	85	-0.75	
	abcd	78	-0.25	

The effects, when plotted, show that the $*$ effects are important, with A and B being the most important. Now suppose that only half of the data had been collected, e.g., the points with •'s.

$$D = -ABC$$

	A	B	C	D
a	+	-	-	-
b	-	+	-	-
c	-	-	+	-
abc	+	+	+	-
d	-	-	-	+
abd	+	+	-	+
acd	+	-	+	+
bcd	-	+	+	+

Yates algorithm:

			Col 1	Col 2	Col 3	Effect	SS
(1)	d	61	122	245	579	($\bar{y}$) 72.38	CF 41905.125
a		61	173	284	-29	-7.25	105.125
b		90	119	-7	97	24.25	1176.125
ab	d	83	165	-22	5	1.25	3.125
c		68	0	51	-11	-2.75	15.125
ac	d	51	-7	46	-15	3.75	28.125
bc	d	85	-17	-7	-5	-1.25	3.125
abc		80	-5	12	19	4.25	36.125

How do we interpret this analysis? What has been lost by our using only one-half the data? Look at the table of $+$ and $-$ below. Observe that effects are confounded.

How do we determine the factors that are confounded? That is, the confounding pattern. First, we decide what factor is to be confounded with D.

For example, $D = -ABC$. This is called the generator. From this we determine the defining relation.

$$DD = -ABCD$$
$$D^2 = -ABCD$$
$$I = -ABCD$$

You can interpret this as $ABCD$ is confounded with the intercept term.

		A	B	C	D	ABC
	(1)	-	-	-	-	-
•	a	+	-	-	-	+
•	b	-	+	-	-	+
	ab	+	+	-	-	-
•	c	-	-	+	-	+
	ac	+	-	+	-	-
	bc	-	+	+	-	-
•	abc	+	+	+	-	+
•	d	-	-	-	+	-
	ad	+	-	-	+	+
	bd	-	+	-	+	+
•	abd	+	+	-	+	-
	cd	-	-	+	+	+
•	acd	+	-	+	+	-
•	bcd	-	+	+	+	-
	abcd	+	+	+	+	+

Next:

$$A \times (-ABCD) = -BCD$$
$$B \times (-ABCD) = -ACD$$
$$AB \times (-ABCD) = -CD$$
$$C \times (-ABCD) = -ABD$$
$$AC \times (-ABCD) = -BD$$
$$BC \times (-ABCD) = -AD$$
$$ABC \times (-ABCD) = -D$$

Therefore A is confounded with -BCD, et cetera. This is called the *alias pattern*. Because A is confounded with -BCD, in the Yates algorithm above the effect nominally estimated for A is in fact an estimate of $A - BCD$. Looking back at the full results, we see that the estimate of the A effect is -8.00 and the BCD estimate is -0.25, while our estimate of $A - BCD$ using half the data is $-8.00 - (-0.75) = -7.25$. IF BCD is only noise, the estimates of A for the full factorial and the half factorial have the same expected value. At any rate, from the half factorial we have estimates of A-BCD, B-ACE, AB-CD, C-ABD, AC-BD, BC-AD, and ABC-D of -7.25, 24.25, 1.25, -2.75, 3.75, -1.25, and 4.25, respectively.

Having examined the alias pattern and decided what it is that you want, how do we decide the treatments that need to be run for this 1/2 fraction of 2^4, or 2^{4-1} fractional factorial.

Set up the table of (+,-) for the first $4-1=3$ factors to be run.

		A	B	AB	C	AC	BC	ABC
(1)	d	-	-	+	-	+	+	-
a		+	-	-	-	-	+	+
b		-	+	-	-	+	-	+
ab	d	+	+	+	-	-	-	-
c		-	-	+	+	-	-	+
ac	d	+	-	-	+	+	-	-
bc	d	-	+	-	+	-	+	-
abc		+	+	+	+	+	+	+

IF $D = -ABC$ ($I = -ABCD$), we match the levels of D with the +, - in the above column. We could also run the other half of the design,

$$I = ABCD,$$

which corresponds to interchanging high and low levels of D. Notice that if we combine the two halves of the design, the design is complete, and, if the two halves constitute blocks, then the blocks are confounded with the defining relation. That is Blocks=ABCD.

We conclude with SAS and Splus code for analyzing an unreplicated 2^4 factorial design; this code can be adapted to the situations discussed in this lecture.

```
options linesize=80  pagesize=100 pageno=1;
***********************************************;
* unrep2n.sas
*
* Construct normal probability plot for
* Unreplicated 2^4 factorial
*
* First homework problem
*
* 9/21/98
************************************************;
 data rate;
   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 @@;
   ab=a*b; ac=a*c; ad=a*d; bc=b*c; bd=b*d; cd=c*d;
   abc=a*b*c; abd=a*b*d; acd=a*c*d; bcd=b*c*d;
   abcd=a*b*c*d;
   output;
   end; end; end; end;
```

```
   cards;
   1.68 1.98 3.28 3.44 4.98 5.70 9.97 9.07 2.07 2.44 4.09 4.53 7.77 9.43 1
 run;

/* if PROC FACTEX is available, then there's an easier way to generate
   the design matrix
*/

proc print;
var a b c d y;
run;

* Compute main effects and interactions and output to a file;

proc reg data=rate outest=regout;
   model y=a b c d ab ac ad bc bd cd abc abd acd bcd abcd;
   title 'Drill rate 2^4 factorial';

proc transpose data=regout out=ploteff name=effect prefix=est;
   var a b c d ab ac ad bc bd cd abc abd acd bcd abcd;
proc print data=ploteff;
   title2 'Dataset produced by the OUTEST option in REG';
   title3 'Transposed to form useful for graphing';

* Compute normal scores ;

proc rank data=ploteff normal=blom out=qqplot;
   var est1;
   ranks normalq;

* Plot normal scores vs the effect estimates;

proc gplot data=qqplot;
   plot normalq*est1=effect;
   title2 'Normal probability plot of effects';

data hnormal;
set ploteff;
heffect=abs(est1);
run;


proc rank data=hnormal  out=hnranks;
   var heffect;
   ranks hneffect;
```

```
run;

data hnormals;
set hnranks nobs=n;
label hneffect='Half Normal score';
hneffect=probit(((hneffect-1/3)/(n+1/3))/2+.5);
run;

proc gplot data=hnormals;
    title2 'Half Normal Probability Plot';
    plot heffect*hneffect=effect;




proc reg data=rate graphics;
   model y=b c d bc cd;
   plot rstudent.*p.  rstudent.*b  rstudent.*c rstudent.*d;
   plot rstudent.*nqq.;
   title2 'Residual plots for model y=b c d bc cd';
run;
```

Here's some Splus code

```
#Get the data and the design in

drill.design<-fac.design(rep(2,4))
y<-c(1.68,1.98,3.28,3.44,4.98,5.70,9.97,9.07,2.07,2.44,4.09,4.53,
     7.77,9.43,11.75,16.3)
drill.df<-data.frame(drill.design,y)
drill.df  #just a check

# Now analyze the data

plot.design(drill.df)  #this gives a first look at the data

drill.aov<-aov(y~(A+B+C+D)^4,drill.df)
coef(drill.aov)
summary(drill.aov)

# Now plot the effects

# a half-normal plot

qqnorm(drill.aov, label=T)
```

```
# a normal plot

qqnorm(drill.aov, full=T, label=T)

#fit a model
drill.aov.model<-aov(y~B+C+D+B*C+C*D, drill.df)
summary(drill.aov.model)

qqnorm(resid(drill.aov.model))
plot(fitted(drill.aov.model),resid(drill.aov.model))
plot(drill.df$B,resid(drill.aov.model))
plot(drill.df$C,resid(drill.aov.model))
plot(drill.df$D,resid(drill.aov.model))
```

Homework: Carefully analyze the data from the helicopter experiment. Analyze the bean data as well. In both cases, bring to class your calculations (including computer code if you used a computer), your plots, and a brief description of your conclusions. I'll collect one of the two analyses. Additionally, finish Chapter 3 of the text and start Chapter 4.

LECTURE 3

Strategy of Experimentation III

Comments:

- Homework

1. Design Resolution

A design is of resolution R if no p factor effect is confounded with any other effect containing less than $R - p$ factors.

For example, a 2^{3-1} design with I=ABC is resolution 3, because

p	$R-p$
1	2

that is main effects are not confounded.

Another example. A 2^{4-1} design with I=ABCD is resolution 4 since

p	$R-p$
1	3
2	2

That is, main effects are not confounded with other main effects or two-factor interactions. Two factor interactions are confounded.

Another example. A 2^{5-1} design with I=ABCDE is resolution 5.

p	$R-p$
1	4
2	3

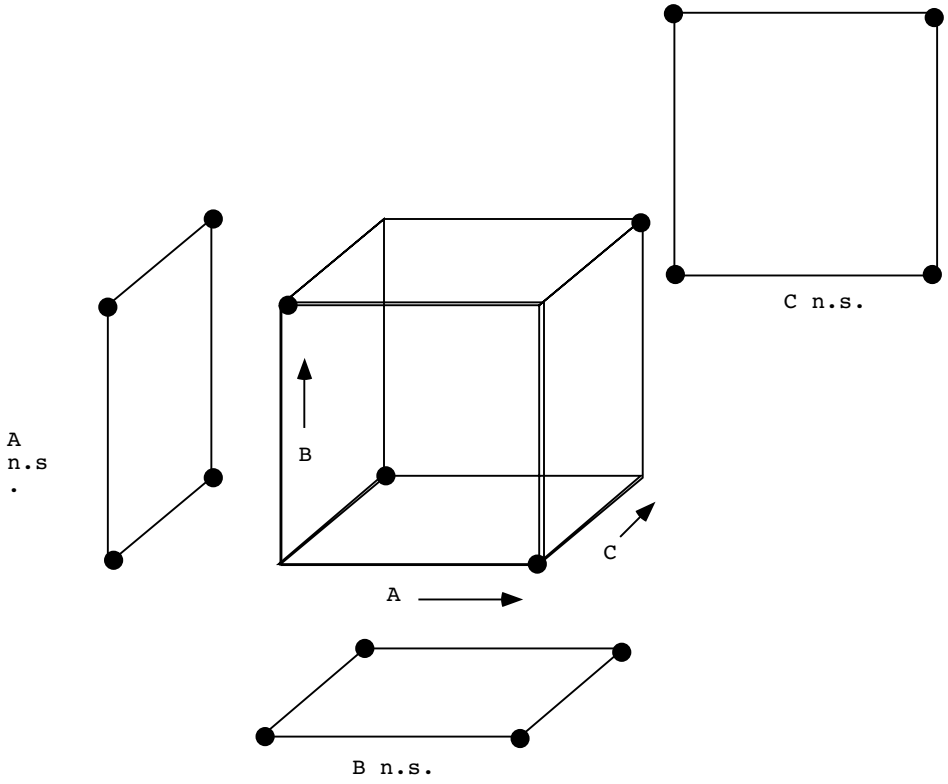
That is, main effects are not confounded with other main effects, two and three-factor interactions. Two-factor interactions are not confounded with other two-factor interaction.

In general, the resolution of the design is the length of the shortest word in the defining relation. The notation commonly used with fractional factorials is 2_R^{K-k}. For example, 2_{IV}^{4-1}.

If, in a particular experimental situation, an experimenter believes that not more than $R - 1$ factors are important, and then uses a resolution R design, and this belief turns out to be true, then the experiment is a complete factorial for the the $R - 1$ factors. To see this, geometrically consider a 2_{III}^{3-1} design with I=ABC.

		A	B	AB
(1)	c	-	-	+
a		+	-	-
b		-	+	-
ab	c	+	+	+

we can project into any dimension and have a 2^2 factorial.



In general, any power of 2 fraction of the design can be used: 1/2, 1/4, 1/8, ...

To see how this works, consider a 1/4 fraction of 2^K, i.e., 2^{K-2}.

Let $K = 5$, $p = 2$, so that $2^{5-2} = 8$ treatment combinations.

Let D=AB and E=AC for the generators. The defining relations are

$$I = ABD$$
$$I = ACE$$

so $I = (ABD)(ACE) = A^2BCDE = BCDE$. This means that we have a resolution 3 design. Remark: the number of defining relations of length R is called the *aberration* of the design. For a given resolution, generally the smaller the aberration the better.

We can find the alias pattern by multiplying the "core" 2^3 factorial by the various relations.

$I = I$	$I = ABD$	$I = ACE$	$I = BCDE$
A	$A(ABD) = BD$	$A(ACE) = CE$	$A(BCDE) = ABCDE$
B	AD	ABCE	CDE
AB	D	BDE	ACDE
C	ABCD	AE	BDE
AC	BCD	E	ABDE
BC	ACD	ABE	DE
ABC	CD	BE	ADE

show treatments run to achieve this pattern

We can use the other 3 possible fractions, e.g.

$$I = -ABD$$
$$I = ACE$$

$$I = ABD$$
$$I = -ACE$$

$$I = -ABD$$
$$I = -ACE$$

References for fractions of two-level factorials:

C. Daniel, *Applications of Statistics to Industrial Experimentation,* Wiley, 1976.

Box and Hunter, "The 2^{K-p} fractional factorial designs," *Technometrics* (1961), pp311– and pp449–.

The original paper on fractional designs is by Finney, "Fractional replication of factorial arrangements," *Annals of Eugenics* **12** (1945) 291–301.

In general, a

$$2^{K-p}$$

fractional factorial has p generators and 2^p defining relations (including I). Each factor is aliased with $2^p - 1$ others, and the design has 2^p distinct fractions. For example,

$p = 1$ 1 generator
2 defining relations
2 fractions

$p = 2$ 2 generators
4 defining relations
4 fractions

If R is the resolution of the design then for a 2^{K-1}, we have $K \geq R$, e.g., ... For a 2^{K-2}, we have $K \geq \frac{3}{2}R$, e.g., ... For a 2^{K-3}, we have $K \geq \frac{7}{4}R$.

In general, if we have a 2^{K-p}_R design then

$$R \leq \frac{2^{p-1}}{2^p - 1} K.$$

Where does this come from? Well, the number of relations that a given letter (=main effect) can appear in is $(1/2)2^p = 2^{p-1}$, and there are K letters. On the other hand, there are $2^p - 1$ relations other than I, and R is the length of the shortest one(s). Thus

$$2^{p-1} K \geq (2^p - 1) R,$$

which gives the relation above. It is possible, however, that a particular design may not be attainable.

1.1. Augmenting fractional designs. After a fractional design has been run, many times the question arises how to supplement the design points. Lots of work has been done on this subject. To give you some ideas along these lines, consider the following situations.

With *resolution III* designs the main effects are confounded with two-factor interactions. Suppose that a factor from the first experiment looks interesting. The factor may be "de-aliased" by running a second fraction in which the sign of the variable has been switched.

For example, suppose we run a 2^{5-2}_{III} design with $D = AB$ and $E = BC$, giving

$$\begin{aligned} I &= ABD \\ I &= BCE \\ I &= (ABD)(BCE) = ACDE \end{aligned}$$

indeed making this resolution III.

The treatment run would be

		A	B	C	D=AB	E=BC
(1)	de	-	-	-	+	+
a	e	+	-	-	-	+
b		-	+	-	-	-
ab	d	+	+	-	+	-
c	d	-	-	+	+	-
ac		+	-	+	-	-
bc	e	-	+	+	-	+
abc	de	+	+	+	+	+

If we found from this experiment that A was important then we could run

		A	B	C	D	E
a	de	+	-	-	+	+
(1)	e	-	-	-	-	+
ab		+	+	-	-	-
b	d	-	+	-	+	-
ac	d	+	-	+	+	-
c		-	-	+	-	-
abc	e	+	+	+	-	+
bc	de	-	+	+	+	+

that is, the first column switched, but everything else unchanged.

If we arrange the 16 treatments for a 2^{5-1}

(1)	e
a	e
b	
ab	
c	
ac	
bc	e
abc	e
d	e
ad	e
bd	
abd	
cd	
acd	
bcd	e
abcd	e

By looking at the $\pm$ table, we find $E = BC$ or $I = BCE$, thus giving 2^{5-1}_{III}. But the alias patterns are

2^{5-2}_{III}

A	BD	ABCE	CDE
B	AD	CE	ABCDE
AB	D	ACE	BCDE
C	ABCD	BE	ADE
AC	BCD	ABE	DE
BC	ACD	E	ABDE
ABC	CD	AE	BDE

2^{5-1}_{III}

A	ABCE
B	CE
AB	ACE
C	BE
AC	ABE
BC	E
ABC	AE
D	BCDE
AD	ABCDE
BD	CDE
ABD	ACDE
CD	BDE
ACD	ABDE
BCD	DE
ABCD	ADE

Note that both designs are resolution 3 but in the second design, A, and two-factor interactions with A, are confounded only with higher-order interactions, giving a resolution V design *for A*. Note that the design could have been resolution V from the beginning.

Fold Over An alternative might occur when it is of interest to free up all the main effects. In this case all you need to do is *switch the signs of all the treatments*. For example, in the previous problem, we would add to the first experimental run

		A	B	C	D=-AB	E=-BC
abc		+	+	+	-	-
bc	d	-	+	+	+	-
ac	de	+	-	+	+	+
c	e	-	-	+	-	+
ab	e	+	+	-	-	+
bc	de	-	+	-	+	+
a	d	+	-	-	+	-
(1)	de	-	-	-	-	-

If we arrange the 16 treatments for a 2^{5-1}

(1)	
a	e
b	
ab	e
c	e
ac	
bc	e
abc	
d	e
ad	
bd	e
abd	
cd	
acd	e
bcd	
abcd	e

By looking at the $\pm$ table for this set of treatment combinations we find that $E = ACD$, giving $I = ACDE$, which is a 2^{5-1}_{IV} design, which means that main effects are *not* confounded with two-factor interactions. Note that two-factor are confounded (including A). The generators for the various fractions are thus

Original Fraction	Enhance A	Fold Over
$D = AB$	$D = -AB$	$D = -AB$
$E = BC$	$E = BC$	$E = -BC$

This type of procedure to create a second fraction is called *folding over*. In general, folding over a resolution III design gives a resolution IV design.

We have seen that, by examining the specific treatments, we can obtain the generators for the defining contrasts. If the design is complicated, this procedure can be messy. It can be done directly with the use of the generators, as is now shown.

In the first pair of designs,

fraction 1 $D = AB$ and $E = BC$, giving, as the generators

$$I = ABD = BCE,$$

the defining relations then begin

$$I = ABD = BCE = ACDE.$$

fraction 2 $D = -AB$ and $E = BC$, giving, as the generators

$$I = -ABD = BCE,$$

the defining relations then begin

$$I = -ABD = BCE = -ACDE.$$

Since $I = BCE$ is common to both fractions, this is the generator when both fractions are put together (this is the result obtained earlier).

In the second pair of designs,

- fraction 1

$$I = ABD = BCE = ACDE$$

- fraction $2D = -AB, E = -BC$, so

$$I = -ABD = -BCE = ACDE.$$

Since I=ACDE is common to both halves the combined fractions have this as the generator.

1.2. More on Augmenting Designs. One problem with the fold-over or enhance-a-single-factor designs is that they require a set of runs as large as the original design. This may or may not be a problem, but there are a few alternatives.

1.2.1. *Adding a few orthogonal runs*. Consider the set of runs above the line.

Run	A	B	C	D=ABC	AB	CD	Block
1	-	-	-	-	+	+	-
2	+	-	-	+	-	-	-
3	-	+	-	+	-	-	-
4	+	+	-	-	+	+	-
5	-	-	+	+	+	+	-
6	+	-	+	-	-	-	-
7	-	+	+	-	-	-	-
8	+	+	+	+	+	+	-
9	+	+	+	+	+	+	+
10	-	-	+	-	+	-	+
11	-	+	+	+	-	+	+
12	+	-	+	-	-	-	+

Suppose that based on the first 8 runs, we decided that A was the strongest effect, but that C and the contrast AB + CD also appeared active. We would like to be able to decide whether AB, CD, or both are actually active. We can do so in 4 runs. For the 4 additional runs (shown below the line), we pick 2 orthogonal vectors (+,+,-,-) and (+,-,+,-). There's only one more orthogonal vector, (+,-,-,+), so we assign it to the largest main effect, A. These choices determine the vector for B. A more or less arbitrary choice for C likewise determines the vector for D. We should also include a blocking effect, because these runs will be performed after the first eight. Finally, since the full matrix is no longer orthogonal, a regression analysis will be needed for the resulting data.

1.2.2. *Optimal Design for Augmenting Designs*. We will talk about optimal design later in the semester, but for now let us suppose that after a first set of runs we have some effects we would like to de-alias. For concreteness, assume as above that A was the strongest effect, but that C and the contrast AB + CD also appeared active. We might then consider the model

$$y = \beta_0 + \beta_A A + \beta_B B + \beta_C D + \beta_D D + \beta_{AB} AB + \beta_{CD} CD + \epsilon.$$

Writing $\mathbf{X}$ as the model matrix, if you've had regression you know that the least squares estimator of the β vector is $\hat{\beta} = (\mathbf{X^t X})^{-1}\mathbf{X^t y}$, with covariance matrix $\sigma^2(\mathbf{X^t X})^{-1}$. This suggests the D-optimality criterion for a design, namely maximizing $|\mathbf{X^t X}|$.

Actually, since we really just want to de-alias AB and CD, we probably want a different criterion. We can write

$$\mathbf{X^t X} = \begin{bmatrix} X_1^t X_1 & X_1^t X_2 \\ X_2^t X_1 & X_2^t X_2 \end{bmatrix},$$

where $\mathbf{X_2}$ would be the model matrix for AB and CD only. Then the lower right submatrix of $\mathbf{X^t X^{-1}}$ is

$$\left(\mathbf{X_2^t X_2} - \mathbf{X_2^t X_1 (X_1^t X_1)^{-1} X_1^t X_2}\right)^{-1}$$

and so the criterion is to maximize

$$\left|\mathbf{X_2^t X_2} - \mathbf{X_2^t X_1 (X_1^t X_1)^{-1} X_1^t X_2}\right|,$$

which is known as the D_s criterion. One might also add the constraint that the two effects have orthogonal contrasts. At any rate, programs such as SAS (using proc optex) can find suitable designs given this optimality condition.

Homework: Work on the two handouts. I will collect one of them.

LECTURE 4

Strategy of Experimentation IV

1. Fractional Factorials Continued

1.1. A Followup Note to the Bean Problem. C. Daniel pointed out that the effect of S, K, and SK are all about the same magnitude, although the sign of S is negative. Suppose we look at a regression model of just these factors,

$$E(y) = \beta_0 + \beta_S S + \beta_K K + \beta_{SK} SK.$$

Simplifying the βs for the moment as $0 -1, 1,$ and $1,$ (and recalling that the settings for each factor are -1 and 1), we have

Treatment	Expected Value
(1)	1
s	-3
k	1
sk	1

This sort of pattern comes up surprisingly often; the regression-interaction model tends to obscure the pattern.

1.2. Resolution III. Let us summarize how designs can be created. To create resolution III designs one assigns the additional factors to the interactions to create the generators. For example, in a

$$2^{7-4},$$

we let $D = AB$, $E = AC$, $F = BC$, $G = ABC$. These designs are called "saturated designs;" with 2^{K-p} runs, one can estimate $2^{K-p} - 1$ main affects, assuming all two-way and higher effects are negligible. If we have few factors, we reduce p and K by equal amounts. That is,

$$2^{6-3}, 2^{5-2}, 2^{4-1}.$$

In each case one fewer generators is needed allowing us more flexability in selecting the confounding.

1.3. Resolution IV. We start by creating, when possible, a resolution III design. Each of these designs may then be "folded over" to create a resolution IV design.

Suppose that we cannot create a resolution III design, e.g., we want to have a 2^{8-4}_{IV}. We would start, using the previous procedure, with a 2^{8-5}_{III} and fold it over. The problem is there is no resolution III design. This can be seen by noting that

there are only 4 interactions but there are 5 extra factors, or simply by noting that with 8 runs one cannot estimate 8 main effects plus an overall mean.

An easier way is to set up the full matrix for $K-p$ variables. Then we confound each extra variable with an interaction with an odd number of letters. For example, a 2^{8-4} has interactions ABC, ABD, ACD, BCD which have an odd number of letters (2 and 4 factor interactions will not work). Therefore, the generators are

$$E = ABC, \quad F = ABD, \quad G = ACD, \quad H = BCD.$$

You may check to determine this is resolution IV by finding all of the defining contrasts (2^4 of them).

1.4. Development of resolution V designs. In this case all main effects and two-factor interactions are not confounded with each other. This means that we must have at least

$$K + \binom{K}{2} = K + \frac{K(K-1)}{2} = \frac{K(K+1)}{2}$$

data points.

Earlier we saw that for resolution V, $K \geq 5$. For example, if $K = 4$, then $4 + 4 * 3/2 = 10$ and $2^{4-1} = 8$, which is not enough data.

For $K = 5$, $5 + 5 * 4/2 = 15$, which means that a 2_V^{5-1} is possible with $E = ABCD$.

For $K = 6, 6 + 6 * 5/2 = 21$, which means that a $2^{6-2} = 16$ will not work. A 2^{6-1} is okay where $F = ABCDE$.

For $K = 7, 7+7*6/2 = 28$. In principle a $2^{7-2} = 2^5 = 32$ might be possible. It does not work out, however (try $F = ABCD$ and $G = BCDE$, for example). Therefore we must use G=ABCDEF for 2^{7-1}.

For $K = 8, 8 + 8 * 7/2 = 36$, while $2^{8-2} = 2^6 = 64$. We wish to pick two five-factor interactions which have minimal overlap. For example,

$$G = ABCD \qquad H = CDEF,$$

or

$$I = ABCDG = CDEFH = ABEFGH.$$

2. Screening Designs

One of the uses of the fractional factorial designs is in "screening" designs. By this we mean that the experimenter has a large number of factors and is interested in determining those that are not important.

In these procedures the experimenter usually must assume that there are no interactions. A little later we will discuss this issue a little further.

In screening designs we wish to expend as little effort as possible. In the first group of these designs we will make them as close to saturated as possible.

2.1. Fractional Designs. Let us start by considering fractional designs of resolution III. Note that we use resolution III because a lower resolution will confound main effects and a higher resolution will require too many treatments to run.

For example, if $K = 7$ we can run $2^{7-4}_{III} = 8$ treatments. They are

	A	B	AB	C	AC	BC	ABC
(1)	-	-	+	-	+	+	-
a	+	-	-	-	-	+	+
b	-	+	-	-	+	-	+
ab	+	+	+	-	-	-	-
c	-	-	+	+	-	-	+
ac	+	-	-	+	+	-	-
bc	-	+	-	+	-	+	-
abc	+	+	+	+	+	+	+

Let the generators be

$$D = AB$$
$$E = AC$$
$$F = BC$$
$$G = ABC$$

giving $I = ABD = ACE = BCF = ABCG$ and the other relations by multiplications giving resolution III.

If $K = 5$, 2^{5-2}_{III}. If $K = 6$, 2^{6-3}_{III}. If $K = 8$, we can't obtain a resolution III design with 8 treatments. Therefore we must use 2^{8-4}, giving 16 treatments which gives a resolution IV design.

We see then, that we are expending $16 - 9 = 7$ more data points than we need. The problem with the fractional designs is that we must have some power of 2 as the number of data points.

2.2. Plackett-Burman Designs. An alternative to these designs, proposed by Plackett and Burman (1946), allows for the number of data points to be a multiple of 4, e.g., $8, 12, 16, \ldots$.

Consider the design for $N = 8$. + + + - + - - is given. The design is then (*show how the design is generated*)

Run							
1	+	+	+	-	+	-	-
2	-	+	+	+	-	+	-
3	-	-	+	+	+	-	+
4	+	-	-	+	+	+	-
5	-	+	-	-	+	+	+
6	+	-	+	-	-	+	+
7	+	+	-	+	-	-	+
8	-	-	-	-	-	-	-

(Note that the final row is added to the end of each design.) Note that with Plackett-Burman designs, if we have fewer factors, say 6 in this case, we drop off columns.

Changing the row order to $1, 2, 3, 6, 5, 7, 4, 8$ gives the 2^{7-4}_{III} we found earlier.

	-G	-B	-C	-D	-A	-E	-F
1	+	+	+	-	+	-	-
2	-	+	+	+	-	+	-
3	-	-	+	+	+	-	+
6	+	-	+	-	-	+	+
5	-	+	-	-	+	+	+
7	+	+	-	+	-	-	+
4	+	-	-	+	+	+	-
8	-	-	-	-	-	-	-

The advantage of the Plackett-Burman designs is that we do not have the power of 2 restriction, the $4N$ being much more flexible. The disadvantage compared to a fractional factorial design, is that the aliasing pattern for a Plackett-Burman design is much more complex. Generally, each main effect is aliased with every 2-way interaction not involving that effect.

Calculations for a Plackett-Burman design are illustrated in the following example. Six flavors were being screened by a small group of judges during the early stages of development of a new food product. The response is the total number of points given to the formulation. Six of the eleven columns of the 12 run Plackett-Burman design were used.

	Flavors						Unused Factors					
Run	A	B	C	D	E	F	G	H	I	J	L	Scores
1	+	+	-	+	+	+	-	-	-	+	-	65
2	-	+	+	-	+	+	+	-	-	-	+	88
3	+	-	+	+	-	+	+	+	-	-	-	52
4	-	+	-	+	+	-	+	+	+	-	-	49
5	-	-	+	-	+	+	-	+	+	+	-	43
6	-	-	-	+	-	+	+	-	+	+	+	52
7	+	-	-	-	+	-	+	+	-	+	+	10
8	+	+	-	-	-	+	-	+	+	-	+	83
9	+	+	+	-	-	-	+	-	+	+	-	69
10	-	+	+	+	-	-	-	+	-	+	+	17
11	+	-	+	+	+	-	-	-	+	-	+	100
12	-	-	-	-	-	-	-	-	-	-	-	18
Contrast	-18	110	-127	-44	268	-10	-52	-14	38	-24	44	
Effect=Contrast/6	-3.0	18.3	-21.2	-7.3	44.7	-1.7	-8.7	-2.3	6.3	-4.0	7.3	

Now using the techniques of QQ plots the above effects may be evaluated.

In addition to the effects being examined by QQ plots, statistical tests can be performed to determine if the effects are statistically significant. In the above example the last 5 columns give effects that are measures of experimental variation. The average effects from these column can be used to estimate the standard deviation of the effects.

$$s = \sqrt{\frac{1}{M}\left(E_G^2 + E_H^2 + \cdots + E_K^2\right)},$$

where E_G is the average effect for column G, etc., and M is the number of contrasts used to assess experimental error.

An effect associated with each of the six flavors is statistically significant if it exceeds $t_{M,.05} \times s$, where t is Student's t based on M d.f.

$$t = 2.57$$
$$s = \sqrt{\frac{1}{5}(-8.7)^2 + \ldots (7.3)^2} = 6.616$$

indicating that B, C, and E are statistically significant.

2.3. Supersaturated Designs. Another stategy for screening designs is to use so-called super-saturated designs. This means that for $K - 1$ factors we have fewer than K data points.

One such group of designs are those proposed by Booth and Cox (1962). Others have been proposed and will be topics at the end of the semester.

In the Booth and Cox designs we have N observations and p parameters, where N is even. Each column of the design matrix will have $N/2$ +1's and $N/2$ −1's. In the Plackett Burman designs, the columns are orthogonal. That is , if $C_i^{n\times 1}$ is the ith column vector, then $C_i' C_j = 0$ for all $i \neq j$. This condition cannot be satisfied for all columns i and j if $K > N - 1$. The rank of the matrix must be $\leq K$. Recognizing this the authors wish to have this requirement satisified as nearly as possible. The criteria that they select is

$$\min \left(\max_{i \neq j} C_i' C_j \right),$$

i.e., find the worst case of dependence and then make that as close to independence as possible.

Furthermore, if two designs have the same value for the above, the design in which the number of pairs of columns attaining the above minimum is chosen.

In the handout, note that their f is our K. In the designs handed out, if K is less than shown, drop off the appropriate number of right hand columns. Also, for design I, for $K = N - 1$, we have the usual Plackett-Burman design. Finally, note that the paper has several more pages of designs.

Hamada and Wu (*JQT*, 1992) have some interesting methods to analyze experiments with complex aliasing. They involve looking at a large number of possible regression models and invoking effect sparsity and the hierarchical principle and enable one to pick up 2nd order interactions even in saturated and supersaturated designs.

2.4. Group Screening Designs. Another approach to the screening problem is the "Group Screening Method." With this procedure the factors are placed in groups, the groups are tested, and then the factors within the significant groups are tested. (This is actually two-stage grouping.)

For example, suppose that $K = 9$ and three groups are formed. In

group X	place	A, B, C
group Y	place	D, E, F
group Z	place	G, H, I

With the 3 groups we can run 2^{3-1}. Let's suppose the half we run is

$$x,\ y,\ z,\ xyz,$$

then with x, for example, actually run

$$A^+,\ B^+,\ C^+,\ D^-,\ E^-,\ F^-,\ G^-,\ H^-,\ I^-.$$

After the experiment is run then an experiment is run on the significant group. For example, if one group is significant, say X, run a 2^{3-1} on A, B, C using $N = 4$. If two groups, say X, Y, are significant, run Plackett Burman, $N = 8$, on A, B, C, D, E, F. This procedure was proposed by Watson (1961) in *Technometrics*. In this paper, the author sets up the following assumptions.

1. All factors have, independently, the same prior probability of being effective.
2. Effective factors have the same effect, $\Delta > 0$.
3. There is no interaction present.
4. The required designs exist.
5. The directions of possible effects are known.
6. Errors are independent, normal with constant variance.
7. $K = gf$, where g is the number of groups, f is the number of factors per group, and K is the number of factors.

In a few minutes we will remove assumption (1).

(2) is required to obtain optimal group size and is not important.

Assumptions (3) and (5) are important because they guarantee that effects can't cancel each other out.

Assumption (4) allows the development of group-sizes that minimize the total number of runs.

Assumption (6) is the usual assumption to do ANOVA.

And assumption (7) allows for equal group sizes.

Using the above assumptions the first question is what is optimum group size. The criteria for "optimum" used by Watson was that the expected number of experimental runs be minimized. Watson showed that

$$E(R) = K(1 - p^K + \frac{1}{f} + \frac{1}{K}),$$

where K equals number of factors, p equals the probability of a factor not being important, f equals the number of factors per group, and R equals the number of runs.

The result is shown in Table 2 (Handout). The first column is prior probability for the factor being important. The second column is the optimum group size. The third column is not important, while the fourth is the probability that a group will contain at least one important factor. The fifth is the probability that a group will contain at leasst two important factors.

As was mentioned earlier the first assumption can be relaxed. What we can do is place the factors into g groups where the factors in the same group have equal probability of being effective but between groups they do not. From the previous handout we can see that important factors should be placed in small groups while unimportant factors should be placed in large groups.

For example, suppose there are 50 factors ($K = 50$) of which 30 have a probability of .02, 10 have a probability of ,07, and 10 have a probability of .12. For

probability =.02, optimum f =8
probability =.07, optimum f =5
probability =.12, optimum f =4

Therefore we might consider

4 groups of 8	32
2 groups of 5	10
2 groups of 4	8
8	50

This gives up 8 groups in the first stage. We could then use a Plackett Burman design with $N = 12$. If we could reduce the number of groups to 7 then we could use 2^{4-1}. For example,

5 groups of 8	40
1 groups of 6	6
1 groups of 4	4
7	50

or

5 groups of 8	40
2 groups of 5	10
7	50

The nearest equal group design would require on of the factors being dropped.

Alternatively, if we decide to go with 12 groups we could consider

4 groups of 6	24
3 groups of 5	15
2 groups of 4	8
1 groups of 3	3
10	50

or

11 groups of 4	44
2 groups of 3	6
13	50

and run a 2^4.

Let's see how this last design might compare to the use of a regular P-B design. Based on the original assumptions, we can expect 2 to 3 factors being effective.

That is,

$$30 \times .02 = .6$$
$$10 \times .07 = .7$$
$$10 \times .12 = .12$$
$$\text{Total} = 2.5 \longleftarrow \text{expected number}$$

Let's say 3, and assume they fall in different groups of size 4. This leaves us with 12 factors to test. We could use a P-B design with $n = 16$. Therefore, the total number of runs is 16(phase I) + 16(phase II) = 32. we could have used a Plackett-Burman design without grouping. This would have required 52 runs, thus saving 20 runs.

These methods have, of course, been extended. One obvious strategy is to have more than two stages.

In conclusion, the most crucial assumptions are that

1. No interaction.
2. The direction of the effects are known.

References

Kleijman, J.P.C., "Screening Designs for Poly Factor Experimentation," *Technometrics*, 1975, pp 487–. A good survey paper.

Plackett, R. L. and Burman, J. P. "The Design of Optimum Multifactorial Experiments," *Biometrika*, 1946, pp 305–325. A basic paper.

John, Peter W. M., *Statistical Design and Analysis of Experiments*, New York: MacMillan, 1971 (recently reprinted by SIAM). Section 9.5 has a brief, high-level explanation of how Plackett-Burman designs were originally generated.

Booth and Cox, "Some systematic supersaturated designs," *Technometrics*, 1962, pp. 489–.

Watson, G. S. "A Study of the Group Screening Methods," *Technometrics*, 1961, pp 371–.

LECTURE 5

Response Surface Methodology I

1. Introduction

Response surface methodology is a collection of experimental strategies, mathematical methods, and statistical inference which enable an experimenter to make efficient empirical exploration of the system of interest.

The work which initially generated interest in the package of techniques was a paper by Box and Wilson in 1951.

Many times these procedures are used to optimize a process. For example, we may wish to maximize yield of a chemical process by controlling temperature, pressure and amount of catalyst.

The basic strategy has four steps:

1. Procedures to move into the optimum region.
2. Behavior of the response in the optimum region.
3. Estimation of the optimum conditions.
4. Verification.

Now let's set up the problem. We have p factors. Call them $x_1, x_2, \ldots, x_p$. We have a response y, and a function ϕ, such that

$$E(y) = \phi(x_1, x_2, \ldots, x_p).$$

Initially, ϕ is usually approximated by a first order regression model over narrow regions of x, that is, where there is little curvature. That is,

$$E(y) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p = \beta_0 + \sum_{i=1}^{p} \beta_i x_i.$$

In regions of higher curvature, especially near the optimum, second order models are commonly used:

$$\begin{aligned} E(y) &= \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p \\ &\quad + \beta_{11} x_1^2 + \cdots + \beta_{pp} x_p^2 \\ &\quad + \beta_{12} x_1 x_2 + \cdots + \beta_{p-1,p} x_{p-1} x_p \\ &= \beta_0 + \sum_{i=1}^{p} \beta_i x_i + \sum_{i=1}^{p} \beta_{ii} x_i^2 + \sum_{i} \sum_{j>i} \beta_{ij} x_i x_j. \end{aligned}$$

Therefore, the overall strategy is to use first order models to "climb" the response surface and then higher order models to sexplore the optimum region.

Let us now consider the first phase of experimentation. There are basically two issues that will be considered. First the types of experimental designs that are used and then procedures to determine where the next experimental design should be run. Remember we are climbing the response surface.

2. First order models

First we will consider designs for fitting first order models. In a regression problem, in matrix notation

$$Y = X\beta + \epsilon,$$

where

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \qquad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix},$$

and X is the design matrix

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix},$$

but we will code the data to center it at 0, and use ± 1.

Using the above coding, one of the designs used with first order models is 2^p factorials. For example, suppose that $p = 3$ and we wish to center the experiment as follows:

$$x_1 = 225$$
$$x_2 = 4.25$$
$$x_3 = 91.5$$

Next we must decide how how far to extend the design from the center. As a rough guideline:

1. Make them far enough apart to allow the effect of the factor to be seen.
2. Make them not so far apart as to feel the surface is curving appreciably.

For example,

$$x_1 \pm 25$$
$$x_2 \pm .25$$
$$x_3 \pm 1.5$$

This gives

x_1	200	250
x_2	4.0	4.5
x_3	90	93

Now let

$$x_1' = \frac{x_1 - 225}{25}$$
$$x_2' = \frac{x_2 - 4.25}{0.25}$$
$$x_3' = \frac{x_3 - 91.5}{1.5},$$

giving

$$X = \begin{bmatrix} 1 & -1 & -1 & -1 \\ 1 & +1 & -1 & -1 \\ 1 & -1 & +1 & -1 \\ 1 & +1 & +1 & -1 \\ 1 & -1 & -1 & +1 \\ 1 & +1 & -1 & +1 \\ 1 & -1 & +1 & +1 \\ 1 & +1 & +1 & +1 \end{bmatrix}.$$

From regression methods we know that

$$\hat{\beta} = (X'X)^{-1}X'y$$
$$\text{Cov}(\hat{\beta}) = \sigma^2(X'X)^{-1}.$$

Notice that for our example

$$(X'X)^{-1} = \begin{bmatrix} 1/8 & 0 & 0 & 0 \\ 0 & 1/8 & 0 & 0 \\ 0 & 0 & 1/8 & 0 \\ 0 & 0 & 0 & 1/8 \end{bmatrix}.$$

Since this is diagonal the estimates of the regression coefficients are independent. Of course we already knew that from the work done earlier in this semester. At any rate,

$$\hat{\beta}_0 = \bar{y}$$
$$\hat{\beta}_i = \frac{\text{effect } i}{2} \qquad i = 1, 2, 3$$
$$\text{Var}(\hat{\beta}_i) = \frac{\sigma^2}{8} \qquad i = 0, 1, 2, 3$$

In evaluating the designs one question we might ask concerns the problem that we might run into if a 2nd order model is required.

Of course, we know that the interactions are orthogonal to the main effects, so they will be okay. The quadratic forms, x_i^2, will give a column of 1's, thus they will be confounded with $\hat{\beta}_0$.

As we have studied earlier we can fractionate the 2-level design. For example, we might have a 2^{3-1} design where we let $C = AB$. The design matrix will be

$$X = \begin{bmatrix} 1 & -1 & -1 & +1 \\ 1 & +1 & -1 & -1 \\ 1 & -1 & +1 & -1 \\ 1 & +1 & +1 & +1 \end{bmatrix}.$$

Again, what happens if we need a quadratic model?

x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	x_1^2	x_2^2	x_3^2
+1	-1	-1	+1	+1	-1	-1	+1	+1	+1
+1	+1	-1	-1	-1	-1	+1	+1	+1	+1
+1	-1	+1	-1	-1	+1	-1	+1	+1	+1
+1	+1	+1	+1	+1	+1	+1	+1	+1	+1

We see

$\hat{\beta}_1$ confounded with $\hat{\beta}_{23}$
$\hat{\beta}_2$ confounded with $\hat{\beta}_{13}$
$\hat{\beta}_3$ confounded with $\hat{\beta}_{12}$
$\hat{\beta}_0$ confounded with $\hat{\beta}_{11}, \hat{\beta}_{22}, \hat{\beta}_{33}$.

The confounding pattern of the interactions we have observed before, of course.

The previous designs do not allow for an estimate of experimental error and therefore do not allow for a test of lack of fit for the model.

For example in the 2^3 design, using the first order model gives

Source	df
Total	7
Regression	3
Residual	4

The "residual" is a composite of both lack of fit and experimental error.

To allow for an estimate of experimental error and a little information about quadratic terms, the 2^p or 2^{p-q} design can be supplemented by n_c center points.

For this design the design matrix is

x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	x_1^2	x_2^2	x_3^2
+1	-1	-1	-1	+1	+1	+1	+1	+1	+1
+1	+1	-1	-1	-1	-1	+1	+1	+1	+1
+1	-1	+1	-1	-1	+1	-1	+1	+1	+1
+1	+1	+1	-1	+1	-1	-1	+1	+1	+1
+1	-1	-1	+1	+1	-1	-1	+1	+1	+1
+1	+1	-1	+1	-1	+1	-1	+1	+1	+1
+1	-1	+1	+1	-1	-1	+1	+1	+1	+1
+1	+1	+1	+1	+1	+1	+1	+1	+1	+1
+1	0	0	0	0	0	0	0	0	0
+1	0	0	0	0	0	0	0	0	0
+1	0	0	0	0	0	0	0	0	0
+1	0	0	0	0	0	0	0	0	0

What we notice is that the quadratic terms can be estimated independently from β_0, although not from each other. Also, as we have run more than one center point, we can obtain an estimate of experimental error. The ANOVA for this design is

Source	df
Total	11
β_1	1
β_2	1
β_3	1
β_{12}	1
β_{13}	1
β_{23}	1
$\beta_{11}, \beta_{22}, \beta_{33}$	1
β_{123}	1
Exp. Error	3

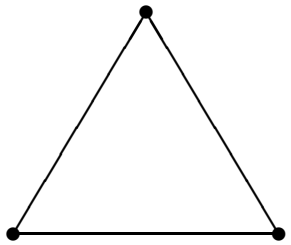
or

Source	df
Total	11
Linear	3
Lack of fit	5
Interaction	4
Quadratic	1
Exp Error	3

While other designs have been proposed, these are the most popular for first order models.

Other first order designs can be used. One type, which we have not seen before, is called a simplex design. This design will be used extensively later on in the semester when we discuss Mixture Experimentation. But we will briefly introduce the concept of a simplex design now.

The simplex designs that we will discuss today are orthogonal designs which have $n = p + 1$ points. Geometrically, the design points represent the vertices of a p-dimensional regular sided figure, or simplex. For example, if $p = 2$, the points form an equilateral triangle.



The design may be constructed with the following procedure. Construct the design matrix,

$$X^{(p+1)\times(p+1)},$$

by letting $X = \sqrt{n}O$, where O is an orthogonal matrix (i.e., $O^{-1} = O^T$). For example, for $p = 2$, we may construct O as follows. First, find a $(p+1) \times (p+1)$ matrix whose columns are independent and whose first column is 1. For example,

$$\begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & -1 \\ 1 & 0 & 2 \end{bmatrix}.$$

Now divide each element of a given column by $\sqrt{\sum c_i^2}$, that is, the length of the column considered as a vector. For our example

$$\begin{aligned} \text{column 1:} \quad & \sqrt{\textstyle\sum c_i^2} = \sqrt{3} \\ \text{column 2:} \quad & \sqrt{\textstyle\sum c_i^2} = \sqrt{2} \\ \text{column 3:} \quad & \sqrt{\textstyle\sum c_i^2} = \sqrt{6} \end{aligned}$$

giving

$$O = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & 0 & 2/\sqrt{6} \end{bmatrix}.$$

Note that $O'O = I$. Thus

$$\begin{aligned} X &= \sqrt{n}O = \sqrt{3}O \\ &= \begin{bmatrix} 1 & \sqrt{\frac{3}{2}} & -\frac{1}{\sqrt{2}} \\ 1 & -\sqrt{\frac{3}{2}} & -\frac{1}{\sqrt{2}} \\ 1 & 0 & \frac{2}{\sqrt{2}} \end{bmatrix}. \end{aligned}$$

Note that

$$X'X = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix},$$

indicating that the design is orthogonal.

If $p = 3$, starting with 2^{3-1},

$$O = \begin{bmatrix} 1/2 & 1/2 & -1/2 & -1/2 \\ 1/2 & -1/2 & -1/2 & 1/2 \\ 1/2 & 1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & 1/2 & -1/2 \end{bmatrix}.$$

Now, $\sqrt{n}O = 2O$, so

$$X = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix}.$$

Notice that this is in fact a 2^{3-1} factorial.

Draw picture.

One final note is that this procedure will not necessarily generate a unique design; this is because different O's can be generated. For example, with $p = 3$,

$$O = \begin{bmatrix} 1/2 & 0 & 1/\sqrt{2} & -1/2 \\ 1/2 & -1/\sqrt{2} & 0 & 1/2 \\ 1/2 & 0 & -1/\sqrt{2} & -1/2 \\ 1/2 & 1/\sqrt{2} & 0 & 1/2 \end{bmatrix},$$

giving

$$X = \begin{bmatrix} 1 & 0 & \sqrt{2} & -1 \\ 1 & -\sqrt{2} & 0 & 1 \\ 1 & 0 & -\sqrt{2} & -1 \\ 1 & \sqrt{2} & 0 & 1 \end{bmatrix}.$$

While both designs are simplex designs one may be preferable to another regarding the biases of the regression coefficients against second order coefficients. As we have seen, with 2^{3-1}

$$\begin{aligned} &\hat{\beta}_0 \text{ is confounded with } \hat{\beta}_{11}, \hat{\beta}_{22}, \hat{\beta}_{33} \\ &\hat{\beta}_1 \text{ is confounded with } \hat{\beta}_{23} \\ &\hat{\beta}_2 \text{ is confounded with } \hat{\beta}_{13} \\ &\hat{\beta}_3 \text{ is confounded with } \hat{\beta}_{12} \end{aligned}$$

In the second design,

x_0	x_1	x_2	x_3	x_1^2	x_2^2	x_3^2	x_1x_2	x_1x_3	x_2x_3
1	0	$\sqrt{2}$	-1	0	2	1	0	0	$-\sqrt{2}$
1	$-\sqrt{2}$	0	1	2	0	1	0	$-\sqrt{2}$	0
1	0	$-\sqrt{2}$	-1	0	2	1	0	0	$\sqrt{2}$
1	$\sqrt{2}$	0	1	2	0	1	0	$\sqrt{2}$	0

This indicates that $x_1^2 + x_2^2 = x_0$ and $x_3^2 = x_0$. Therefore,

$$\begin{aligned} &\hat{\beta}_0 \text{ is confounded with } \hat{\beta}_{11}, \hat{\beta}_{22}, \hat{\beta}_{33} \\ &\hat{\beta}_1 \text{ is confounded with } \hat{\beta}_{13} \\ &\hat{\beta}_2 \text{ is confounded with } \hat{\beta}_{23} \\ &\hat{\beta}_3 \text{ is confounded with } \hat{\beta}_{11} \text{ and } \hat{\beta}_{22} \end{aligned}$$

Now let's briefly discuss how the computations can be carried out for these simplex designs. Continuing with the example suppose we used 2^{3-1}

(1)	x_1	x_2	$x_3 = x_1x_2$	y	
	+1	-1	-1	+1	51.6
	+1	+1	-1	-1	54.1
	+1	-1	+1	-1	31.2
	+1	+1	+1	+1	51.6

You can see that $X'X = \text{diag}(4, 4, 4, 4)$, and since

$$\hat{\beta} = (X'X)^{-1}X'y,$$

we have

$$\begin{aligned} \hat{\beta}_0 &= \bar{y} = \frac{195.5}{4} = 48.9 \\ \hat{\beta}_1 &= \frac{112.7 - 82.8}{4} = 7.5 \\ \hat{\beta}_2 &= \frac{89.8 - 105.7}{4} = -4.0 \\ \hat{\beta}_3 &= \frac{110.2 - 85.3}{4} = 6.2 \end{aligned}$$

This gives us the estimated regression model

$$\hat{y} = 48.9 + 7.5x_1 - 4.0x_2 + 6.2x_3.$$

The next issue is one of obtaining a graphical representation of the model. A procedure that has been found to be successful is one of using response contours. For two variables,

insert picture

This may be obtained from the estimated regression model

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2,$$

by solving for, say, x_2,

$$x_2 = \frac{\hat{y} - \hat{\beta}_0}{\hat{\beta}_2} - \frac{\hat{\beta}_1}{\hat{\beta}_2} x_1.$$

Now all we need to do is select a value of $\hat{y}$, say $\hat{y}_0$, and we have a line

$$x_2 = \frac{\hat{y}_0 - \hat{\beta}_0}{\hat{\beta}_2} - \frac{\hat{\beta}_1}{\hat{\beta}_2} x_1.$$

This can be repeated for several values of $\hat{y}$ giving the contour plot.

LECTURE 6

Response Surface Methodology II

References:
The basic paper:

Box, G. E. P. and K. B. Wilson (1951). "On the experimental attainment of optimal conditions," *JRSS, Series B*, **13** 1–.
Reviews:

Hill, William J. and William G. Hunter (1966). "A review of rsponse surface methodology: a literature survey," *Technometrics*, 571–591.

Mead, R., and D. J. Pike (1975). "A review of response surface methodology from a biometric point of view," *Biometrics*, 803–.

Myers, Khori, and Carter (1989). "Response surface methodology 1966–88," *Technometrics*, 66–88.

There are several good books as well.

1. Steepest Ascent

We know that to maximize the response, the movement of the design center must be in the direction of the directional derivatives of the response function, that is, in the direction of

$$\frac{\partial \phi}{\partial x} = (\frac{\partial \phi}{\partial x_1}, \ldots, \frac{\partial \phi}{\partial x_p}).$$

We then multiply by a constant A so that

$$\Delta x = A \frac{\partial \phi}{\partial x},$$

where

$$A = \frac{r}{\sqrt{\sum \left(\frac{\partial \phi}{\partial x_i}\right)^2}}.$$

Thus $\sum \Delta x_i^2 = r^2$.

For the first order model,

$$\frac{\partial \phi}{\partial x_i} = \hat{\beta}_i,$$

so $\Delta x_i = A\hat{\beta}_i$ and $A = r/\sqrt{\sum \hat{\beta}_i^2}$. From this we see that the movement of x_i up the path of steepest ascent is proportional to $\hat{\beta}_i$. Since this is the case it is easier

not to pick particular values of r but rather fix a value of $\hat{\beta}_i$ and make the other changes proportional to it. To see the procedure consider the following example:

		x_1	x_2	x_3	x_1'	x_2'	x_3'
1	Base Level (Center Point)	0	0	0	225	4.25	91.5
2	Unit Change (divisor in codings)	1	1	1	25	0.25	1.5
3	$\hat{\beta}$ from coded regression model	7.5	-4.0	6.2			
4	Uncoding of slope to original x scale				$7.5 \times 25 = 187.5$	-1.0	9.3
5	Change relative to one unit change in $\hat{\beta}_1 = 1$	1	-0.53	0.83	1	$\frac{-1.0}{187.5}$=-0.0053	0.0496
6	Path of steepest ascent	1	-0.53	0.83	250	4.12	92.7
		2	-1.07	1.65	275	3.98	94.0
		3	-1.6	2.48	300	3.85	95.2

Now discuss where the next design might be centered. Probably would choose center set of points to center design.

Where to center the next experiment?

1. Center the design on the path of steepest ascent.
2. Remember that points outside of ± 1 are extrapolations from the model.
3. One possibility is to try a set of points along the path of steepest ascent, and then center the next experiment where the largest response was.
4. A more conservative approach would be to center the next experiment near the boundary of the first experiment.

This procedure would be continued until is appears that a stationary point has been reached. One should look at lack of fit if at all possible as this might indicate a region of high curvature.

2. Second Order Models

Having reached the region the optimum we wish to explore this region more carefully, using a second order model. Orthogonality now becomes less important, while the prediction variance $\text{Var}(\hat{y})$ becomes more important. Again, there are two issues, one of design and the other of analysis.

2.1. Design. Since we are using second order models we must have at least three levels for each factor. For example, a 3^k design would be possible. For example, if $K = 2$,

x_1	x_2
-1	-1
0	-1
+1	-1
-1	0
0	0
+1	0
-1	+1
0	+1
+1	+1

We use a full second order model of the form

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11}(x_1^2 - c) + \beta_{22}(x_2^2 - c),$$

where $c = \bar{x_i^2} = 2/3$ for 3^K designs. *Discuss why we subtract off* c. The design matrix is

x_0	x_1	x_2	x_1x_2	$x_1^2 - 2/3$	$x_2^2 - 2/3$
1	-1	-1	+1	1/3	1/3
1	0	-1	0	-2/3	1/3
1	+1	-1	-1	1/3	1/3
1	-1	0	0	1/3	-2/3
1	0	0	0	-2/3	-2/3
1	+1	0	0	1/3	-2/3
1	-1	+1	-1	1/3	1/3
1	0	+1	0	-2/3	1/3
1	+1	+1	+1	1/3	1/3

giving

$$X'X = \begin{bmatrix} x_0 & x_1 & x_2 & x_1x_2 & x_1^2 - c & x_2^2 - c \\ 9 & & & & & 0 \\ & 6 & & & & \\ & & 6 & & & \\ & & & 4 & & \\ & & & & 2 & \\ 0 & & & & & 2 \end{bmatrix}$$

and $\text{Cov}(\hat{\beta}) = \sigma^2 (X'X)^{-1}$. We see that the estimates are independent of each other.

The problem with 3^k models is the large number of data points for relatively small k (e.g., $k = 3$ gives $n = 27$, and $k = 4$ gives $n = 81$). In addition,$\text{Var}(\hat{\beta}_{ii})$ is relatively large;

$$\text{Var}(\hat{\beta}_{ii}) = \frac{\sigma^2}{2} \qquad \text{for } k = 2$$

as opposed to

$$\text{Var}(\hat{\beta}_i) = \frac{\sigma^2}{6} \qquad \text{for } k = 2.$$

2.2. Central Composite Designs. Another type of design, developed by Box and Wilson (1951), is called the Central Composite Design.

These designs are first order designs (2^k or a fraction of 2^k, but almost always resolution V or better) augmented by center points and star, or axial, points. For example, for $k = 2$,

$$\begin{aligned} 2^k &= 2^2 = 4 \\ \text{center point(s)} &= 1 \\ \text{star points} &= 4 \end{aligned}$$

giving, as the design matrix

x_1	x_2
-1	-1
+1	-1
-1	+1
+1	+1
0	0
$-\alpha$	0
$+\alpha$	0
0	$-\alpha$
0	$+\alpha$

Pictorially (assuming $\alpha > 1$)
Draw picture

Again using the same model as before we will let

$$c = \frac{2^k + 2\alpha^2}{n} = \bar{x_i^2},$$

where n equals the total number of points in the design.

For the full second order model the design matrix will be, for $k = 2$, so $c = (4 + 2\alpha^2)/9$, and if we let $\alpha = 2$, $c = (4 + 8)/9 = 4/3$,

x_0	x_1	x_2	x_1x_2	$x_1^2 - 4/3$	$x_2^2 - 4/3$
1	-1	-1	+1	-1/3	-1/3
1	+1	-1	-1	-1/3	-1/3
1	-1	+1	-1	-1/3	-1/3
1	+1	+1	+1	-1/3	-1/3
1	0	0	0	-4/3	-4/3
1	-2	0	0	8/3	-4/3
1	+2	0	0	8/3	-4/3
1	0	-2	0	-4/3	8/3
1	0	+2	0	-4/3	8/3

giving

$$X'X = \begin{bmatrix} x_0 & x_1 & x_2 & x_1x_2 & x_1^2 & x_2^2 \\ 9 & 0 & 0 & 0 & 0 & 0 \\ 0 & 12 & 0 & 0 & 0 & 0 \\ 0 & 0 & 12 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 20 & -12 \\ 0 & 0 & 0 & 0 & -12 & 20 \end{bmatrix}.$$

There are several points worth making

1. The linear and quadratic terms have greater precision than those of 3^k designs (the interaction is slightly worse off).
2. The quadratic terms have a non-zero covariance.
3. For $k = 2$, the number of experimental runs is the same. In general, it is much less for the central composite.

k	CC	3^k
2	9	9
3	15	27
4	25	81

The question to discuss is that of selecting α. One of the criteria that can be used to select α is to make the estimates of the quadratic terms orthogonal, that is, make the $X'X$ matrix diagonal.

If we use a 2^{k-p} design with $2k$ axial points and n_c center points then

$$\alpha = \left\{ \frac{\left[\left(2^{k-p} + 2k + n_c\right)^{\frac{1}{2}} - 2^{\frac{k-p}{2}} \right]^2 2^{k-p}}{4} \right\}^{\frac{1}{4}},$$

will make $X'X$ diagonal. For example, if we let $p = 0$ and $n_c = 1$,

k	α
2	1.000
3	1.216
4	1.414
5	1.596
6	1.761

Now a logical question that might be asked is whether 3^k or the orthogonal central composite design is better. An answer requires criteria to be developed to compare the designs. Later on in the semester we will have a discussion on optimality criteria. At that time we will develop procedures which will allow us to compare these designs.

Another criteria that is used to select α is to make the design *rotatable*. A design is said to be rotatable when the variance of $\hat{y}$ is a function only of the distance from the center of the design and not a function of the direction.

Geometrically

Draw picture

Points 1 and 2 are both distance ρ from center $(0, 0)$ and so would have the same $\mathrm{Var}(\hat{y})$. The concept of rotatable is not uniquely related to second order models or central composite designs. Myers, and Box and Hunter give a detailed description relating to rotatibility.

Remember from regression that

$$\hat{y} = x^{1\times p}\hat{\beta}^{p\times 1},$$

where $\mathrm{Var}\,\hat{\beta} = \sigma^2(X'X)^{-1}$. Thus

$$\mathrm{Var}(\hat{y}) = \sigma^2 x^{1\times p}(X'X)^{-1}x',$$

where $x^{1\times p}$ is a particular row of the design matrix.

For example, it turns out that, for first order designs, an orthogonal design is also a rotatable design. As an illustration consider a 2^2 factorial

$$X = \begin{bmatrix} 1 & -1 & -1 \\ 1 & +1 & -1 \\ 1 & -1 & +1 \\ 1 & +1 & +1 \end{bmatrix}$$

$$X'X = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1/4 & 0 \\ 0 & 0 & 1/4 \end{bmatrix}$$

If we take x to be row 1 of X, then

$$x(X'X)^{-1}x' = \begin{bmatrix} 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1/4 & 0 \\ 0 & 0 & 1/4 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}$$

$$= \begin{bmatrix} 1/4 & -1/4 & -1/4 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} = 3/4$$

For row 2,

$$x(X'X)^{-1}x' = \begin{bmatrix} 1/4 & 1/4 & -1/4 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = 3/4$$

and so on for the remaining two rows.

It must be emphasized that in general, for other than first order models, orthogonal designs are not necessarily rotatable and vice versa.

For central composite designs, the design can be made rotatable if

$$\alpha = 2^{(k-p)/4}.$$

k	2^{k-p}	p	star	center	n	α
2	4	0	4	1	9	1.414
3	8	0	6	1	15	1.682
4	16	0	8	1	25	2.000
5	16	1	10	1	27	2.000
6	32	1	12	1	45	2.378
5	32	0	10	1	43	2.378
6	64	0	12	1	77	2.828

For $p = 2$

Draw picture

all points lie on a circle (except center point) $x_1^2 + x_2^2 = 2$. The design matrix is

$$X = \begin{bmatrix} 1 & -1 & -1 & +1 & 1/9 & 1/9 \\ 1 & +1 & -1 & -1 & 1/9 & 1/9 \\ 1 & -1 & +1 & -1 & 1/9 & 1/9 \\ 1 & +1 & +1 & +1 & 1/9 & 1/9 \\ 1 & -\sqrt{2} & 0 & 0 & 10/9 & -8/9 \\ 1 & +\sqrt{2} & 0 & 0 & 10/9 & -8/9 \\ 1 & 0 & -\sqrt{2} & 0 & -8/9 & 10/9 \\ 1 & 0 & +\sqrt{2} & 0 & -8/9 & 10/9 \\ 1 & 0 & 0 & 0 & -8/9 & -8/9 \end{bmatrix}.$$

$$X'X = \begin{bmatrix} 9 & & & & & 0 \\ & 8 & & & & \\ & & 8 & & & \\ & & & 4 & & \\ & & & & 44/9 & -28/9 \\ 0 & & & & -28/9 & 44/9 \end{bmatrix}.$$

$$(X'X)^{-1} = \begin{bmatrix} 1/9 & & & & & 0 \\ & 1/8 & & & & \\ & & 1/8 & & & \\ & & & 1/4 & & \\ & & & & .34375 & .21875 \\ & & & & .21875 & .34375 \end{bmatrix}.$$

This gives for the first row of the design matrix

$$\begin{aligned} \mathrm{Var}(\hat{y}) &= \sigma^2 x (X'X)^{-1} x' \\ &= \sigma^2 \begin{bmatrix} 1/9 & -1/8 & -1/8 & 1/4 & .0625 & .0625 \end{bmatrix} x' \\ &= \sigma^2 [1/9 + 2/8 + 1/4 + .0139] \\ &= 0.625\sigma^2. \end{aligned}$$

For one of the axial points,

$$\begin{aligned}\operatorname{Var}(\hat{y}) &= \sigma^2 x(X'X)^{-1}x' \\ &= \sigma^2 \begin{bmatrix} 1/9 & -\sqrt{2}/8 & 0 & 0 & .1875 & -.0625 \end{bmatrix} x' \\ &= \sigma^2 [1/9 + 2/8 + .2639] \\ &= 0.625\sigma^2.\end{aligned}$$

They will be equal for all points in the design except for the center point. (Recall that all of the points are equidistant from the center.)

For some of the rotatable CCD (when fractions are used) the points do not all lie on a hypersphere. The designs are still rotatable as the points that are equidistant from the center will have equal variance.

In the above designs for second order models we have noted that they all have a single center point. As with the first order designs, in order to obtain an estimate of experimental error, multiple center points are recommended.

In general, what conditions are required for a design to be rotatable? (See Box and Hunter for a much more complete discussion of this topic.)

For a linear model, the criterion that must be met is (assuming that $\sum_{u=1}^{n} x_{iu} = 0$ and $\sum_{u=1}^{n} x_{iu}^2 = n$) that

$$\sum_{u=1}^{n} x_{iu}x_{ju} = 0 \qquad i \neq j = 1, \ldots, p.$$

This means that

$$X'X = \begin{bmatrix} n & & & 0 \\ & n & & \\ & & \ddots & \\ 0 & & & n \end{bmatrix}.$$

This yields the 2^k type of design. Also the simplex designs that we discussed are rotatable.

For a second order model the criteria are (again assuming $\sum_{u=1}^{n} x_{iu} = 0$)

$$\sum_{u=1}^{n} x_{iu}x_{ju} = 0 \qquad i \neq j = 1, \ldots, k$$

$$\sum_{u=1}^{n} x_{iu}x_{ju}x_{ku} = 0 \qquad i, j, k = 1, \ldots, k$$

$$\sum_{u=1}^{n} x_{iu}^4 = 3\sum_{u=1}^{n} x_{iu}^2 x_{ju}^2, \qquad i \neq j$$

This means that all off diagonal elements, except for the pure quadratics, are equal to zero.

For example, if we have a central composite design, with axial points at $\pm\alpha$, then

$$\sum x_{iu}^4 = 2^{k-p} + 2\alpha^4$$
$$\sum x_{iu}^2 x_{ju}^2 = 2^{k-p} \qquad i \neq j$$

giving $2^{k-p} + 2\alpha^4 = 3(2^{k-p})$ or $\alpha = 2^{(k-p)/4}$, a result that we saw earlier.

While central composite designs are the most often used class of rotatable designs, nonetheless many other types of rotable designs exist, especially for k fairly small. For example, if $k = 2$, then points lying on a circle and equidistant, along with a center point, forms a rotatable design. For example, if we place six points on a circle

Draw picture

Notice that we have divided the 360^o into six sections, $60, 120, \ldots, 300$. If we let the distance from the center to each point (the radius) be equal to 1, then

x_1	x_2
1	0
0.5	$\sqrt{.75}$
-0.5	$\sqrt{.75}$
-1	0
-0.5	$-\sqrt{.75}$
.5	$-\sqrt{.75}$

Note that $\sum x_{1u} = \sum x_{2u} = 0, \sum x_{1u}x_{2u} = 0, \sum x_{1u}^4 = \sum x_{2u}^4 = 2.25, \sum x_{1u}^2 x_{2u}^2 = 0.75$, and so $\sum x_{1u}^4 = \sum x_{2u}^4 = 3\sum x_{1u}^2 x_{2u}^2$.

Note that if we don't add the center point then the design matrix will be singular

$$X = \begin{bmatrix} x_0 & x_1 & x_2 & x_1^2 & x_2^2 x_1 x_2 & \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & .5 & \sqrt{.75} & .25 & .25 & \sqrt{.75}/2 \\ 1 & -.5 & \sqrt{.75} & .25 & .75 & -\sqrt{.75}/2 \\ 1 & -1 & 0 & 1 & 0 & 0 \\ 1 & -.5 & -\sqrt{.75} & .25 & .75 & \sqrt{.75}/2 \\ 1 & .5 & -\sqrt{.75} & .25 & .75 & -\sqrt{.75}/2 \end{bmatrix}.$$

Note that $x_1^2 + x_2^2 = x_0$. If we add the center point the confounding disappears.

The question that arises is how many center points. One proposal, made by Box and Hunter, is to have roughly the same precision as elsewhere in the design. To achieve near uniform precision the number of center points needed is show below.

k	2^{k-p}	number of center points
2	4	5
3	8	6
4	16	7
5	16	6
5	32	10
6	32	9
6	64	15

In general most researchers will not run this many points at the center simply to achieve homogeneous precision.

In general, for spherical designs, 3–5 center points will be sufficient. For face-centered cube designs, that is, CCD with $\alpha = 1$, 1 or 2 center points will do.
References: Box, G.E.P. and J. S. Hunter (1957). "Multifactor experimental designs for exploring response surfaces," *Ann. Math. Stat.* pp. 195–.

2.3. Box-Behnken Designs. Reference: Box, G. E. P., and D. W. Behnken (1960). *Technometrics*, pp. 455–.

Box Behnken designs are fractional 3^k factorials. The designs either meet, or approximately meet, the criterion of rotatability. These designs are formed by combining two-level factorial designs with incomplete block designs.

For example, $k = 3$,

	T_1	T_2	T_3
b_1	$*$	$*$	
b_2	$*$		$*$
b_3		$*$	$*$

.

For each line, replace "$*$" with a 2^2 factorial and put 0 in for blank cells. This gives

x_1	x_2	x_3
-1	-1	0
+1	-1	0
-1	+1	0
+1	+1	0
-1	0	-1
+1	0	-1
-1	0	+1
+1	0	+1
0	-1	-1
0	+1	-1
0	-1	+1
0	+1	+1

Then add the indicated number of center points, in this case, 3.
Draw picture.
For other values of k, the paper shows how the designs can be constructed.

2.4. An Example. Before we go any further, let us consider an example. Suppose that a first order model is initially used, namely a 2^3 design.

$$X = \begin{bmatrix} 1 & -1 & -1 & -1 \\ 1 & +1 & -1 & -1 \\ 1 & -1 & +1 & -1 \\ 1 & +1 & +1 & -1 \\ 1 & -1 & -1 & +1 \\ 1 & +1 & -1 & +1 \\ 1 & -1 & +1 & +1 \\ 1 & +1 & +1 & +1 \end{bmatrix}$$

$$Y = \begin{bmatrix} 55.9 \\ 70.6 \\ 67.5 \\ 68.6 \\ 63.3 \\ 68.0 \\ 68.8 \\ 62.4 \end{bmatrix}$$

$$X'X = \begin{bmatrix} 8 & & & 0 \\ & 8 & & \\ & & 8 & \\ 0 & & & 8 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 525.1 \\ 14.1 \\ 9.5 \\ -0.1 \end{bmatrix}$$

$$\hat{\beta} = (X'X)^{-1}X'Y = \begin{bmatrix} 65.6375 \\ 1.7625 \\ 1.1875 \\ -0.0125 \end{bmatrix}$$

$$Y'Y = 34629.47$$

$$\hat{\beta}'X'Y = 34502.385$$

$$\text{Resid}_{SS} = 127.0856$$

$$\text{Resid}_{MS} = 31.7712$$

From an examination of the model and the data there is an indication that we have reached a stationary point. We therefore would like to supplement this design so that a second order model can be estimated. *Brief discussion of blocking—more later.*

x_1	x_2	x_3	y
0	0	0	66.9
+2	0	0	65.4
-2	0	0	56.9
0	+2	0	67.5
0	-2	0	65.0
0	0	+2	68.9
0	0	-2	60.3

It should be noted that this design is neither orthogonal nor rotatable, although it is a central composite design. If we fit a full second order model we get

$$X'X = \begin{bmatrix} x_0 & x_1 & x_2 & x_3 & x_1^2 & x_2^2 & x_3^2 & x_1x_2 & x_1x_3 & x_2x_3 \\ 15 & & & & 16 & 16 & 16 & & & 0 \\ & 16 & & & & & & & & \\ & & 16 & & & & & & & \\ & & & 16 & & & & & & \\ & & & & 40 & 8 & 8 & & & \\ & & & & 8 & 40 & 8 & & & \\ & & & & 8 & 8 & 40 & & & \\ & & & & & & & 8 & & \\ & & & & & & & & 8 & \\ 0 & & & & & & & & & 8 \end{bmatrix}.$$

What we notice is that the pure quadratic terms are correlated (the correlation with $\hat{\beta}_0$ can be removed if we subtract $\sum x_i^2$ from the pure quadratic terms). While this design is not rotatable the correlation among the quadratic terms is characteristic of that design also.

We find the estimated regression model to be

$$\begin{aligned}\hat{y} = 67.71 &+ 1.944x_1 + 0.906x_2 + 1.096x_3 \\ &- 1.539x_1^2 - 0.264x_2^2 - 0.676x_3^2 \\ &- 3.088x_1x_2 - 2.188x_1x_3 - 1.212x_2x_3\end{aligned}$$

The ANOVA table relating to this analysis is

Source	df	SS=ms	df	ms
Total		14279.33		
β_1	1	60.45		
β_2	1	13.14	3	30.62
β_3	1	18.28		
β_{12}	1	76.26		
β_{13}	1	38.28	3	42.10
β_{23}	1	11.76		
β_{11}	1	31.24		
β_{22}	1	0.62	3	12.31
β_{33}	1	5.07		
Residual	5	4.85		

LECTURE 7

Response Surface Methodology III

1. Canonical Form of Response Surface Models

To examine the estimated regression model we have several choices. First, we could plot response contours. Remember that we set $\hat{y}$ to some specified value, y_0, and trace out contours relating x_1, x_2, and x_3.

An alternative is to reduce the equation to its "canonical form." That is, we form an equation of the form

$$y - y_s = \lambda_1 w_1^2 + \lambda_2 w_2^2 + \lambda_3 w_3^2,$$

where y_s is the center of the contours (that is, the stationary point) and w_1, w_2, and w_3 are a new set of axes called the principal axes. The coefficients λ_1, λ_2, and λ_3 give the shape of the surface (they are the eigenvalues of a matrix to be defined shortly).

Using matrix notation for a bit, we could write the model as

$$\hat{y} = \hat{\beta}_0 + X'\hat{\beta} + X'\hat{B}X,$$

where here X is just the linear part of the design matrix, β is the vector of linear coefficients, and

$$B = \begin{bmatrix} \hat{\beta}_{11} & \hat{\beta}_{12}/2 & \dots & \hat{\beta}_{1k}/2 \\ \vdots & \hat{\beta}_{22} & \dots & \hat{\beta}_{2k}/2 \\ & & \ddots & \vdots \\ \hat{\beta}_{1k}/2 & \dots & \hat{\beta}_{kk} & \end{bmatrix}.$$

To find the stationary point X_s, differentiate $\hat{y}$ to find

$$\frac{\partial y}{\partial X} = \hat{\beta} + 2\hat{B}X.$$

Setting this to zero, we find

$$X_s = -\hat{B}^{-1}\hat{\beta}/2.$$

As to finding the right form for the λ's and the z's, you might recall that if you form a matrix M with columns equal to the normalized eigenvectors of $\hat{B}$, then

$$M'\hat{B}M = \Lambda,$$

where Λ is a diagonal matrix with diagonal elements equal to the eigenvalues of $\hat{B}$. Now write

$$\begin{aligned} Z &= X - X_s \\ W &= M'Z, \end{aligned}$$

we have

$$\begin{aligned} \hat{y} &= \hat{\beta}_0 + X'\hat{\beta} + X'\hat{B}X \\ &= \hat{\beta}_0 + (Z + X_s)'\hat{\beta} + (Z + X_s)'\hat{B}(Z + X_s) \\ &= [\hat{\beta}_0 + X_s'\hat{\beta} + X_s'\hat{B}X_s] + Z'\hat{\beta} + Z'\hat{B}Z + 2X_s'\hat{B}Z \\ &= \hat{y}_s + Z'\hat{B}Z, \end{aligned}$$

because $2X_s'\hat{B}Z = -Z'\hat{\beta}$ from the definition of X_s. Rotating the coordinate system, we have

$$\begin{aligned} \hat{y} &= \hat{y}_s + Z'\hat{B}Z \\ &= \hat{y}_x + W'M'\hat{B}MW \\ &= \hat{y}_s + M'\Lambda M, \end{aligned}$$

which is what we want.

The reason to do this is because the eigenvalues, the diagonal values of Λ, can tell a great deal about the stable point.

- If the eigenvalues are all negative, the stable point is a maximum.
- If the eigenvalues are all positive, the stable point is a minimum.
- If the eigenvalues are of mixed sign, the stable point is a saddle.

That's not all. The relative sizes of the eigenvalues also tell a great deal. For example, if most of the eigenvalues are large positive numbers but a few are near zero, then there is a ridge in the graph of the response function. Moving along that ridge will make little difference in the value of the response (but might make a big difference in some other aspect of the system, like cost, for example).

Let's illustrate this with an example. Say

$$\hat{y} = 81.22 + 1.97x_1 + 0.22x_2 - 3.93x_1^2 - 1.38x_2^2 - 2.22x_1x_2.$$

To find y_s, we must find the values of x_1 and x_2 which represent the stationary point.

$$\begin{aligned} \frac{\partial \hat{y}}{\partial x_1} &= 1.97 - 7.86x_1 - 2.22x_2 = 0 \\ \frac{\partial \hat{y}}{\partial x_2} &= 0.22 - 2.76x_2 - 2.22x_1 = 0 \end{aligned}$$

Solving this system of equations gives

$$\begin{aligned} x_{1,s} &= 0.30 \\ x_{2,s} &= -0.16, \end{aligned}$$

which in turn gives

$$\hat{y}_s = 81.49.$$

To find the eigenvalues of B, we solve

$$\begin{aligned}
0 &= |B - \lambda I| \\
&= \begin{vmatrix} \hat{\beta}_{11} - \lambda & \hat{\beta}_{12}/2 \\ \hat{\beta}_{12}/2 & \hat{\beta}_{22} - \lambda \end{vmatrix} \\
&= \begin{vmatrix} -3.93 - \lambda & -1.11 \\ -1.11 & -1.38 - \lambda \end{vmatrix} \\
&= (-3.93 - \lambda)(-1.3\lambda) - (-1.11)(-1.11) \\
&= \lambda^2 + 5.31\lambda + 4.19
\end{aligned}$$

This yields

$$\lambda = \frac{-5.31 \pm \sqrt{5.31^2 - 4(1)(4.19)}}{2} = \frac{-5.31 \pm 3.38}{2}$$

or

$$\begin{aligned}
\lambda_1 &= -4.35 \\
\lambda_2 &= -0.96.
\end{aligned}$$

Note that the choice of λ_1 or λ_2 is not important.

The next step, following the determination the eigenvalues, is finding the eigenvectors.

Graphically, what we are doing is (in two dimensions)
Draw picture.
w_1 and w_2 are the major and minor axes. The eigenvectors corresponding to each λ_i can give

$$z_i = m_{i1}(x_1 - x_{1,s}) + m_{i2}(x_2 - x_{2,s}),$$

which shows how the axes are translated. If we let

$$M = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} = \begin{bmatrix} M_1 & M_2 \end{bmatrix},$$

then the M_i are the eigenvectors determined by

$$[B - \lambda I]M_i = 0.$$

For example, for $\lambda_1 = -4.35$,

$$\begin{bmatrix} -3.93 + 4.35 & -1.11 \\ -1.11 & -1.38 + 4.35 \end{bmatrix} \begin{bmatrix} m_{11} \\ m_{21} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

or

$$\begin{aligned}
0.42m_{11} - 1.11m_{21} &= 0 \\
-1.11m_{11} + 2.97m_{21} &= 0.
\end{aligned}$$

Note that there is no unique solution to this system of equations, because λ_1 was chosen to make the matrix of coefficients singular.

What we want is to find a proportional relationship, subject to the constraint that $m_{11}^2 + m_{21}^2 = 1$. For example, setting $m'_{11} = 1$, we find $m'_{21} = .42/1.11 = 1.11/2.97 = .3784$. Since $\sqrt{1^2 + .3784^2} = 1.07$, we have $m_{11} = 1/1.07 = 0.94$ and $m_{21} = .3784/1.07 = 0.35$. Similarly, we could find that $m_{21} = 0.35$ and $m_{22} = -0.94$, so

$$M = \begin{bmatrix} .94 & .35 \\ .35 & -.94 \end{bmatrix}.$$

Note that the eigenvectors are orthogonal.

At any rate, we now know that

$$w_1 = .94(x_1 - .30) + .35(x_2 + .16)$$
$$w_2 = .35(x_1 - .30) - .94(x_2 + .16).$$

Draw picture

As mentioned earlier, when all eigenvalues are negative the stationary point is a (local) maximum. When they are not all equal, the function looks like an ellipsoid near the maximum. In our case, the ellipses of the contour plots are elongated along the w_2 axis. This means a small value in w_1 corresponds to a larger value for w_2 as far as giving the same value of $\hat{y} - y_s$.

Having found the canonical form we can now find how it can be used to describe the surface without having to plot the response contours. First, $(x_{1,s}, x_{2,s})$ is the stationary point with y_s being the response at the stationary point. The eigenvalues λ_1, λ_2 give the behavior of the response as we move away from the stationary point.

In our example, $\lambda_1 = -4.35$ and $\lambda_2 = -.96$, which means that the response decreases as we move away from the stationary point. Furthermore, since

$$|\lambda_1| > |\lambda_2|$$

we know that the contours are elongated along the w_2 axis.

If $\lambda_1 = \lambda_2$ then the ellipses are actually into circles.

If λ_1 and λ_2 are both positive then the response is minimized at the stationary point.

If λ_2 is close to zero then we have a stationary ridge. This means that we have a variety of x_1 and x_2 which will result in the maximum value.

In the above situations, the optimum conditions have been reached. All that remains is that we run a experiment in the region the verify the model.

If $\lambda_1 < 0$ and $\lambda_2 > 0$ then we have found a stationary point that is a saddle point. For example

Draw picture

. We have reached a maximum on one axis and a minimum along the other axis. If our objective was to maximize the response then our next experiment should be in the direction of w_2.

If λ_1 and λ_2 are both negative but the stationary point lies well outside the region of the design

Draw picture

This is called a rising ridge. It means we have not yet found the proper region of the maximum, a common occurance. Again the next experiment needs to be run up the ridge.

2. Blocking in Response Surface Designs

The last issue that we will discuss is that of blocking response surface designs. We have already discussed blocking of two-level designs, which can be used in first order designs.

Now we will discuss blocking of second order models. The important issue is to determine how, if at all, one can assign treatments to blocks so that the block effects will be orthogonal to the model coefficients.

First we must consider the conditions necessary for the parameter estimates to be orthogonal to blocks. Let z_{mu} be a dummy variable which takes value $+1$ if the u-th point is in the m-the block and zero otherwise. The model is then

$$Y_u = \beta_0 + \sum_{i=1}^{k} \beta_i x_{iu} + \sum_{i=1}^{k} \beta_{ii} x_{iu}^2 + \sum_{i=1}^{k} \sum_{j>i} \beta_{ij} x_{iu} x_{ju} + \sum_{m=1}^{b} \delta_m (z_{mu} - \bar{z}_m) + \epsilon_u.$$

To have orthogonality of parameter estimates and block effects, we need

$$\sum_{u=1}^{N} x_{iu}(z_{mu} - \bar{z}_m) = 0 \qquad \text{for } i \le k \text{ and all } m$$

$$\sum_{u=1}^{N} x_{iu}^2(z_{mu} - \bar{z}_m) = 0 \qquad \text{for } i \le k \text{ and all } m$$

$$\sum_{u=1}^{N} x_{iu} x_{ju}(z_{mu} - \bar{z}_m) = 0 \qquad \text{for } i \ne j \le k \text{ and all } m$$

With the usual coding for x's so that $\sum x_{iu} = 0$ and $\sum x_{iu} x_{ju} = 0$ for $i \ne j \le k$, these conditions are equivalent to

$$\sum_{i=1}^{N} x_{iu} z_{mu} = 0 \qquad \text{for } i \le k \text{ and all } m$$

$$\sum_{u=1}^{N} x_{iu}^2 z_{mu} = \sum_{u=1}^{N} x_{iu}^2 \bar{z}_m \qquad \text{for } i \le k \text{ and all } m$$

$$\sum_{u=1}^{N} x_{iu} x_{ju} z_{mu} \qquad \text{for } i \ne j \le k \text{ and all } m.$$

The first equation implies that the sum of the observations in the mth block, for variable x_i, is zero. The third equation implies that the cross-product of x_i and x_j sums to 0 in the mth block.

In the second equation, $\bar{z}_m$ is the proportion of observations that occur in block m. Therefore, this equation specifies that the contributions from block m to the total SS for each variable x_i is proportional to the number of runs in the block.

For example, consider the following design.

x_1	x_2	x_3	
1	1	1	
1	-1	-1	
-1	1	-1	
-1	-1	1	Block 1
0	0	0	
0	0	0	
1	1	-1	
1	-1	1	
-1	1	1	Block 2
-1	-1	-1	
0	0	0	
0	0	0	
-1.633	0	0	
1.633	0	0	
0	-1.633	0	
0	1.633	0	Block 3
0	0	-1.633	
0	0	1.633	
0	0	0	
0	0	0	

As for the conditions

- The first equation: $\sum x_{iu} = 0$ within each block.
- The third equation: $\sum x_{iu}x_{ju} = 0$ within each block.
- The second equation. Now $\sum_{u=1}^{N} x_{iu}^2 = 13.33$, the total SS for each variable. For block 1, the sum $\sum_{u=1}^{N} x_{iu}^2 z_{mu} = 4$. For block 2, the sum is 4. For block 3, the sum is 5.33. On the other hand, for block 1, $\sum x_{iu}^2 \bar{z}_m = 4$. For blocks 2 and 3, the sums are 4 and 5.33

Since the conditions are met, the blocking in this design will be orthogonal to the parameter estimates.

In general, for a central composite design, the question that remains is what value of α do we select in order to make the blocks orthogonal to the regression parameters.

First, let's say we want two blocks. In this case we will place the 2^p points in one block and the axial points in the second block. In addition, we'll place c_F center points in the factorial block and C_A center points in the axial block.

From the second condition,

$$\frac{N_A}{N} SS_{TOT} = SS_A, \qquad \frac{N_F}{N} SS_{TOT} = SS_F,$$

which implies that

$$\frac{N_A}{N_F} = \frac{SS_A}{SS_F}.$$

In turn, this means

$$\frac{\sum_{\text{axial block}} x_{iu}^2}{\sum_{\text{factorial block}} x_{iu}^2} = \frac{\text{number of points in axial block}}{\text{number of points in factorial block}} = \frac{2k + C_A}{2^k + C_F}.$$

The left hand side of this equation is simply $2\alpha^2/2^k$, so

$$\frac{2\alpha^2}{2^k} = \frac{2k + C_A}{2^k + C_F},$$

yielding

$$\alpha = \sqrt{\frac{2^k(2k + C_A)}{2(2^k + C_F)}}.$$

Thus is the experimenter requires two blocks, the value of α given by this equation, for the specified values of C_A and C_F, gives a Central Composite Design that blocks orthogonally. For example, if $k = 3$ and $C_A = C_F = 2$, then

$$\alpha = \sqrt{\frac{2^3(2 \cdot 3 + 2)}{2(2^3 + 2)}} \approx 1.7889.$$

What about rotatability? We know that that is also a desirable criterion. We have seen that for rotatability,

$$\alpha = 2^{k/4}.$$

This implies that, for the design to be both rotatable and blocked orthogonally,

$$2^{k/4} = \sqrt{\frac{2^k(2k + C_A)}{2(2^k + C_F)}}$$

or

$$2^{k/2} = \frac{2^k(2k + C_A)}{2(2^k + C_F)}.$$

Thus, for a given value of k, the question remains of finding an appropriate value for C_A and C_F to make the design rotatable. For example, if $k = 2$, we find

$$2 = \frac{4(4 + C_A)}{2(4 + C_F)},$$

or $C_F = C_A$.

If $k = 3$, we get $1.5 + C_A = .5625C_F$, making no solution possible. That doesn't mean, however, that you might not pick a design with orthogonal blocking and near-rotatability. For example, with $C_A = 2$ and $C_F = 3$, for orthogonal blocking we need $\alpha = 1.7056$. For rotatability, we would want $\alpha = 1.6818$. By going with $C_A = 2, C_F = 3$, and $\alpha = 1.7056$, we will get orthogonal blocking and near-rotatability.

For $k = 4$, we get $C_F = 2C_A$.

Some experimental situations dictate the need for more than 2 blocks. To achieve this we fractionate the 2-level portion of the design. We have discussed

how we can fractionate this portion of the design earlier. Again we need to determine the appropriate value for α.

$$\alpha = \sqrt{\frac{2^{k-p}(2k + C_A)}{2(2^{k-p} + C_F)}}.$$

For example, if $C_A = C_F = 2, k = 3$, and $p = 1$, we have

$$\alpha = \sqrt{\frac{2^{3-1}(2 \cdot 3 + 2)}{2(2^{3-1} + 2)}} \approx 1.7889,$$

the design that we saw earlier.

LECTURE 8

Response Surface Methodology IV

1. Bias and Variance

If y_x is the response of the system at the point x, or in short hand, $y_x = f(x)$, then we can write $\eta_x = E(y_x)$. This is the true, and unknown model. We approximate η_x with $\hat{y}_x$, for example $\hat{y}_x = \hat{\beta}_0 + \hat{\beta}_1 x$, but we need to always remember that such a model is at best a workable approximation.

To measure how well out design will do, we could look at the mean squared error

$$\begin{aligned} E(\hat{y}_x - \eta_x)^2 &= E\left(\hat{y}_x - E\hat{y}_x) + (\hat{y}_x - \eta_x)\right)^2 \\ &= E(\hat{y}_x - E(\hat{y}_x))^2 + (E(\hat{y}_x - \eta_x))^2. \end{aligned}$$

To normalize this for changes in the number of design points, N, and for different values of σ, we can multiply each term by N/σ^2, and write the result as $M_X = V_X + B_X$.

The question is how the design affects M_X. Without loss of generality, let us assume that $\sum X_i = 0$, so that $\sigma_X^2 = \sum X_i^2/N$. We have

$$\begin{aligned} \mathrm{Var}(\hat{y}_X) &= \mathrm{Var}(x\hat{\beta}) \\ &= \mathrm{Var}(x(X'X)^{-1}X'y) \\ &= x(X'X)^{-1}X'\,\mathrm{Var}(y)X(X'X)^{-1}x' \\ &= \sigma^2 x(X'X)^{-1}x'. \end{aligned}$$

Thus $\mathrm{Var}(\hat{y}_X) = \sigma^2(\frac{1}{N} + \frac{x^2}{N\sigma_X^2})$, for the simple linear regression model outlined above (remember $x' = [1x]$.) Thus $V_X = 1 + \frac{x^2}{\sigma_X^2}$.

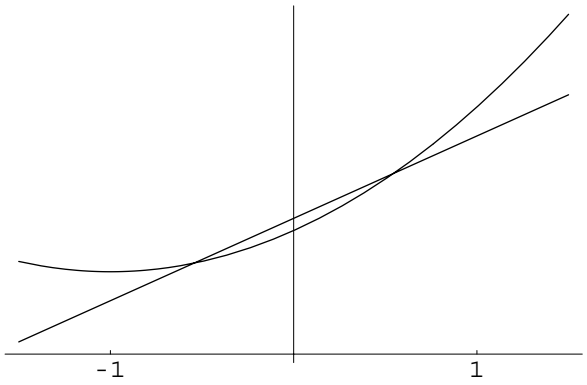
Suppose we have a 3 point symmetric design $(-x_0, 0, x_0)$, so that $\sigma_X^2 = \frac{\sum X_i^2}{N} = 2x_0^2/3$, and

$$V_X = 1 + \frac{3x^2}{2x_0^2},$$

with a design space of $X \in [-4/3, 4/3]$. Now suppose f is actually quadratic, so that

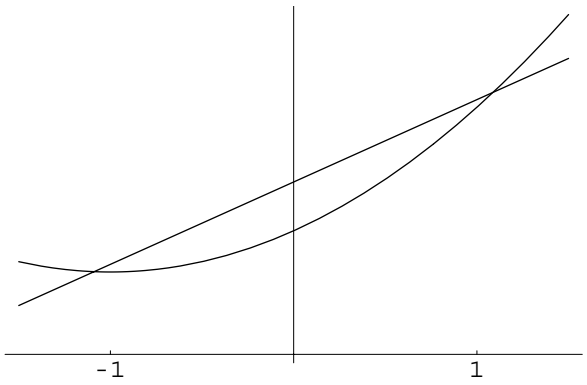
$$y = \beta_0 + \beta_1 x_1 + \beta_{11} x_1^2 + \epsilon.$$

Consider $x_0 = 2/3$.

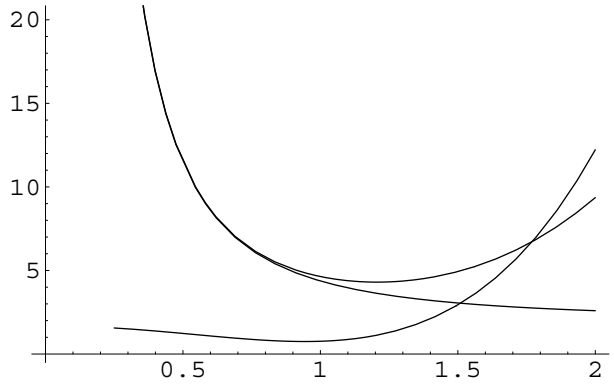


Here we have smaller bias, but bigger variance. In fact, $V_X = 1 + 27x^2/8$.

Consider $x_0 = 4/3$.



This has bigger bias and smaller variance. Here $X_X = 1 + 27x^2/32$. This is better if you are confident about your first-order model.



In fact, the minimum of $M = V + B$ is attained at $x_0 = 1.2$. Without the center point, the minimum would be closer in.

We write the p-th moment of the design as $m_p = \sum X_I^p / N$, and let μ_p be the moments of a (symmetric) prior. We can then write

$$\begin{aligned} M &= V + B \\ &= (1 + \frac{\mu_2}{m_2}) + \alpha^2 \left((m_2 - \mu_2)^2 + (\mu_4 - \mu_2^2) + \frac{m_3^2 \mu_2}{m_2^2} \right), \end{aligned}$$

where $\alpha = \sqrt{N} \frac{\beta_{11}}{\sigma}$. By using a symmetric design, $m_3 = 0$, so the last term drops out. To reduce bias, set $m_2 = \mu_2$. To reduce variance, set m_2 large.

We can thus distinguish among

1. the all variance case, $1 + \frac{\mu_2}{m_2}$, where we want to maximize m_2,
2. the all-bias case, where we want $m_2 = \mu_2$, and
3. the balanced case, which we can index with $V/(V + B)$.

		$\sqrt{m_2/\mu_2}$	
$V/(V+B)$		Uniform prior	Normal prior
All bias	0	1.00	1.00
	.2	1.02	1.04
	.33	1.04	1.09
$V = B$	.5	1.08	1.16
	.67	1.14	1.26
	.8	1.25	1.41
All variance	1	∞	∞

Let us briefly consider designs that minimize squared bias. Suppose a polynomial model of degree d_1,

$$\hat{y}(x) = x_1' b_1,$$

is fitted to the data, while the true model is a polynomial of degree d_2,

$$\eta(x) = x_1' \beta_1 + x_2' \beta_2.$$

If we write

$$M_{11} = \frac{X_1' X}{N},$$
$$M_{12} = \frac{X_1' X_2}{N},$$

and similarly μ_{11} and μ_{12} for the moments under the prior on the design region of interest, it can be shown that, whatever the values of β_1 and β_2, a necessary and suficient condition for the squared bias to be minimized is that

$$M_{11}^{-1} M_{12} = \mu_{11}^{-1} \mu_{12}.$$

One way to satisfy this, of course, is to match to design moments with the prior moments.

As an example, consider a minimum bias design for the k-sphere, where we are going to assume a linear model, but a quadratic model actually holds. If we take a uniform prior over the interior of the sphere (meaning all points inside the sphere are of equal interest), the prior is thus constant within the sphere and zero outside. Since $d_1 = 1$ and $d_2 = 2$, we will be interested only in moments up to

$d_1+d_2 = 3$. Because of the symmetry of the sphere, only the pure second moments will be non-zero. The second moments will be $1/(k+2)$.

These conditions are satisfied by any two-level fractional factorial design of resolution IV if it is scaled.

For example, with $k = 3$, the 2^3 factorial with points placed at $\pm a$ will have second moment $1/N \sum x_{iu}^2 = a^2$. This will equal $1/(k+2)$ if we set $a = 1/\sqrt{5} \approx .447$, or place the vertices on a sphere of radius $\sqrt{3/5} \approx .7746$.

References:

Box & Draper, A basis for the selection of a response surface design, *J. Amer. Stat. Assoc.*, **54**, 1959, 622–654.

Box & Draper, The choice of a second order rotatable design, *Biometrika*, **50**, 1963, 335–352.

2. Practical Summary

Over the last eight weeks we have discussed how to procedd from ignorance to optimization bliss. It might be worth a quick review of how to put it all together.

1. Begin with a screening design to pick out active factors. A resolution III design is all that's really needed. Typically one uses a fractional factorial (but beware of making $k - p$ too small to have a good background against which to pick out the active factors) or Plackett-Burman design.
2. If there are any ambiguities from the screening runs, consider running additional experiments to resolve them. This could be a fold-over design, an enhance-one-factor design, or just a few runs chosen on an *ad hoc* basis. You will almost certainly want to analyze the two sets of runs together, but remember to include a blocking term in your model.
3. Now that you have decided on the active factors, re-analyze your results using just those factors. From the parameter estimators, find the direction of steepest ascent.
4. Decide how you will move in the direction of steepest ascent. One approach is to select a set of runs along the path of steepest ascent for the next design. A more conservative approach is to perform another full or fractional factorial design centered along the path of steepest ascent. The braver you are, the farther out that center can be. If the first design was spread out over a wide region, you might pick the next design to be a smaller size; you might also wish to re-scale based on the relative sizes of the active effects as found from the screening design.
5. Based on the results of the last design, you should have a good idea of where to place your next design. You will probably want to include a pair of center points to check for curvature. If you have plenty of runs still available, and if you have any uncertainty about your choice of active factors, you could do a small fractional factorial design to check that. You could include just the factors that you think are unimportant, or those plus the weakest active factors, or everything. If you don't pick up any new information at least you'll have a good estimate of σ^2.

6. Once you detect curvature, it's time to use a second order design. If you used a (fractional) factorial design with center points, with an appropriate choice of α you can augment that design with center and axial points to get a second-order design with parameter estimates orthogonal to any block effects. Some canonical analysis and ridge analysis will tell you where to move next.
7. If you need another second order design, if possible do it all at once rather than in blocks. That will save you a few center runs. You could also consider a central composite design where the factorial part is just resolution III*. In general, because second order designs use a lot of design points and thus cost a lot of money, you do want to be sure of your design before you implement it.
8. Once you are sure of your optimum point, use a few runs to verify it.

3. EVOP

The techniques that we have been examining for the past few weeks are designed to optimize a process in the Research and Development situation. However, these techniques can also be used in a process already in production. This technique is called *Evolutionary Operation*, or EVOP. Basically, the idea is to let the production facility act as a laboratory. Unlike the laboratory, the production facility cannot make large changes in settings, nor would one expect to find large effects (since presumably there has already been some optimization of the process). On the other hand, it is easier to collect a large amount of data, since the plant is running anyway. A large amount of data will mean smaller standard errors for the effect estimates, so even small effects may be readily detected. In effect, we are looking for smaller things in the grass, but we are cutting the grass shorter.

To understand the basic idea, consider the following diagram showing the possible evolution of a species of lobster. It is supposed that a particular mutation produces a type of lobster with "length of claws" and "pressure attainable between claws" corresponding to a point P in the diagram.

Draw picture. horiz axis, length of claws, vert axis, pressure between claws, contours show percent surviving long enough to reproduce in a given environment.

If the dots indicate offspring, then those that have the greatest chance of survival will help move the scatter of points toward the optimum. The key components of the process of natural selection is

1. Variation
2. Selection of "favorable" variants.

It is this strategy that EVOP tries to exploit. That is, in the EVOP method a carefully planned cycle of minor variants on the process is agreed upon. The routine of plant operation then consists of running each of the variants in turn and continually repeating the cycle. In this way we use routine manufacture to generate not only the product we require but also the information we need to improve it.

The basic strategy is to use small designed experiments, while the process is running, to determine the optimum conditions. Since the improvements may come

in small amounts, it may be necessary to have several replicates in order to improve the "signal to noise ratio" to the point of detecting the improvements.

It is generally found that, while there may be many potentially important independent variables, no more than two or three of them are examined at a time. The reason is that, in practice, this seems to be the limit that can be handled on a routine basis.

The types of designs which are commonly used with EVOP are 2^2 or 2^3 designs, sometimes supplemented with a center point. Of course, it sometimes becomes necessary to block the designs.

With these designs, linear models are commonly used, possibly with interaction terms added in, to approximate the response surface. Then paths of steepest ascent, or some other procedure is used to determine the next region of experimentation.

Simplex design for EVOP.

Consider the following example of EVOP. Suppose there is a chemical process running in your plant for which the EVOP committee would like to vary temperature t and reaction time r. The current operating co nditions are $t = 150^oC$ and $r = 30$ minutes. The EVOP procedure will use a 2^2 design with a center point, run in the sequence (150,30), (145,73), (155,75), (155,74), (145,72). This run order will be easy for plant personnel to remember, and will confound any time or nuisance factor effects with blocks. The cycle is run at least twice. In our example, suppose the means for the various settings are 73, 72, 75.5, 74.5, 72.5. This gives a temperature effect of 2.75, a time effect of .75, an interaction effect of .25, and a "change-in-mean effect" of .50. This last is a measure of curvature, computed as $(1) - ((2) + (3) + (4) + (5))/4$.

Comments on EVOP

- Selection of variables
- Excuse of large run-to-run variation
- EVOP versus SPC
- Training and personnel

EVOP references

The basic paper: Box, G. E. P. (1957), "Evolutionary operation: a method for increasing industrial productivity," *Applied Statistics*, pp 81–101.

A review article: Hunter, William G., Kittrell, J. R. (1966) "Evolutionary operation: A review," *Technometrics*, **8**, pp. 389–397.

A survey article: Hahn, Gerald J., Dershowitz, Arthur F. (1974) "Evolutionary operation today – Some survey results and observations," *Appl Stat,* **23** , pp. 214–218.

A book: Box, G. E. P. and N. R. Draper (1969). *Evolutionary Operation*, New York: Wiley.

LECTURE 9

Mixture Designs

1. Design and Analysis of Mixture Experiments

There are many times when the product we are interested in is a mixture. In other words, we are more interested in the proportions than the total amounts of the components. In product formulations, examples would be gasoline, soaps or detergents, beverages, cake mixes, soups, and so on. There are examples in process engineering, as well. For example, in the production of semiconductor wafers we might be interested in the proportions of various acids for the acid wash.

The fact that the proportions must add up to one is the key attribute of mixture designs. Specifically, the settings for various factors must satisfy

$$x_i \geq 0, \qquad \text{for all } i$$
$$\sum_i x_i = 1$$

The design region for mixture proportions is a simplex, a regularly sided figure of dimenstion $k-1$ with k vertices (and usually embedded into a k dimensional space. For example, with two factors, the simplex is the line segment from (0,1) to (1,0). With three factors, the simplex would have vertices at (1,0,0), (0,1,0), and (0,0,1). There is a corresponding simplex coordinate system.
Draw pictures.

We can now consider models for mixture experiments. The usual first order model is

$$E(y) = \beta_0 + \sum \beta_i x_i.$$

However, since $\sum x_i = 1$ for a mixture model, the β_i's will not be uniquely determined. We could choose to eliminate one of the x_i's, but a better approach was suggested by Scheffé. In the equation above multiply β_0 by $1 = \sum x_i$ to get

$$E(y) = \sum (\beta_0 + \beta_i) x_i.$$

Relabeling the β_i's, we get the following canonical forms.
Linear:

$$E(y) = \sum \beta_i x_i,$$

Quadratic:

$$E(y) = \sum \beta_i x_i + \sum \sum_{i<j} x_i x_j,$$

Special Cubic:

$$E(y) = \sum \beta_i x_i + \sum \sum_{i<j} x_i x_j + \sum \sum \sum_{i<j<k} x_i x_j x_k,$$

Full Cubic:

$$E(y) = \sum \beta_i x_i + \sum \sum_{i<j} \beta_{ij} x_i x_j + \sum \sum_{i<j} \delta_{ij} x_i x_j (x_i - x_j) + \sum \sum \sum_{i<j<k} x_i x_j x_k.$$

This last is rarely used. There are many other possible models. We mention just one, the Draper and St. John model, which can be useful when some of the components work well in small amounts (spices, for example, in food products).

$$E(y) = \sum \beta_i x_i + \sum \nu_i x_i^{-1}$$
$$E(y) = \sum \beta_i x_i + \sum \sum_{i<j} \beta_{ij} x_i x_j + \sum \nu_i x_i^{-1}.$$

The terms in the canonical mixture models have simple interpretations.
Draw pictures.
Geometrically, the parameter β_j represents the expected response from a pure mixture with $x_j = 1$ (and all other components zero). The $\sum \beta_i x_i$ term is called the linear blending term. The quadratic terms should not be thought of as interaction but instead are called nonlinear blending terms. If β_{ij} is positive, the term is synergistic, while if it is negative it is called antagonistic.

For constructing an ANOVA table, the usual formulas apply. That is

$$SS_{Total} = \sum (y_i - \bar{y})^2$$
$$SS_{Reg} = \sum (\hat{y}_i - \bar{y})^2$$
$$SS_{Error} = (y_i - \hat{y}_i)^2.$$

Of course, the first has $n - 1$ degrees of freedom, the second $p - 1$, and the last $n - p$. The F statistic is, as usual,

$$F = \frac{SS_{Reg}/(p-1)}{SS_{Error}/(n-p)},$$

, while

$$R^2 = \frac{SS_{Reg}}{SS_{Total}},$$

and

$$R^2_{Adjust} = 1 - \frac{SS_{Error}/(n-p)}{SS_{Total}/(n-1)},$$

Let's look at a simple, though real, example courtesy of Lynne Hare. First, the data.

Stearine	Oil	SFI-50F
1	0	14.7
2/3	1/3	17.5
1/3	2/3	24.0
0	1	35.5

First, let's fit a linear model $E(y) = \beta_1 x_1 + \beta_2 x_2$. The results are

ANOVA

Source	df	SS	MS	F
Total	3	256.37	85.46	
Model	1	237.36	237.36	24.96
Residual	2	19.01	9.51	

$$R^2 = .926$$
$$R_A^2 = .889$$

Now we will try a quadratic model $E(y) = \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$. The results are

ANOVA

Source	df	SS	MS	F
Total	3	256.37	85.46	
Model	2	256.27	128.14	1281.4
Residual	1	0.10	0.10	

$$R^2 = .9996$$
$$R_A^2 = .9988$$
$$\text{Additional SS due to curvature} = 18.91$$
$$F = 189.1$$
$$F_{.05}(1, 1) = 161.4$$

There are a number of standard mixture designs. The first set are known as the Simplex-Lattice $\{q, m\}$ designs, due to Scheffé. Here we have q components with $m + 1$ equally spaced values from 0 to 1. The Simplex-Lattice Design includes every possible combination of these (remembering that $\sum x_i = 1$). The total number of points is then

$$\binom{m+q-1}{m}.$$

For example, here are three 3-component examples: $\{3, 2\}$, $\{3, 3\}$, and $\{3, 4\}$. *Draw picture*.

Another popular design is the Simplex-Centroid, also due to Scheffé. This will consist of $2^q - 1$ points. There are

- q vertices of the form $(1, 0, \ldots, 0)$
- $\binom{q}{2}$ points of the form $(1/2, 1/2, 0, \ldots, 0)$
-
-
- $\binom{q}{r}$ points of the form $(1/r, 1/r, \ldots, 0)$
- 1 point of the form $(1/q, \ldots, 1/q)$.

Here's the picture for the 4-component example.

For testing lack-of-fit, a popular choice of design is a simplex-centroid with q added interior points of the form $((q+1)/2q, 1/2q, 1/2q, \ldots, 1/2q)$. These points are sometimes called "axial check points" Here's a picture of a 3-component, 10 run design.

As an example, consider measurements of SFI-50 resulting from blends of stearine, vegetable oil solids.

Run	Stearine	Veg. Oil	Veg. Oil Solids	SFI-50 F
1	1	0	0	4.6
2	0	1	0	35.5
3	0	0	1	55.5
4	1/2	1/2	0	14.4
5	1/2	0	1/2	25.7
6	0	1/2	1/2	46.1
7	1/3	1/3	1/3	27.4
8	2/3	1/6	1/6	14.5
9	1/6	2/3	1/6	32.0
10	1/6	1/6	2/3	42.5

With this data, the x_2x_3 term is insignificant, giving a model of

$$y = 4.6x_1 + 35.9x_2 + 56.0x_3 - 21.5x_1x_2 - 16.6x_1x_3.$$

With this model, we get $R_A^2 = .9981$, and the ANOVA table looks like

ANOVA

Source	df	SS	MS
Total	9	2250.58	
Model	4	2248.02	562.00
Residual	5	2.57	.51

The additional SS due to x_1x_2 and x_1x_3 is 37.06, giving an F-statistic of 36.33 and a p-value of 0.0011.

Although a program like Design-Expert will analyze mixture designs with aplomb, analyzing the results with SAS is a little tricky. See the code below; the first proc glm will calculate the estimates correctly, but not the ANOVA, F-test, R^2, and so on (because the no-intercept option does not adjust the SS for the overall mean); the second proc glm will calculate the ANOVA correctly, but not the estimates of the linear terms, but the additional code will calculate the correct linear estimates. The data concerns an experiment with fruit juice. One final comment about data analysis: any residual checking should be done with Studentized residuals, because points in mixture designs can have substantial differences in their leverage values. (Recall that the Studentized residual r_i is given by

$$r_i = \frac{y_i - \hat{y}_i}{\sqrt{\hat{\sigma}^2(1 - h_{ii})}},$$

where h_{ii} comes from the hat matrix $H = X(X'X)^{-1}X'$.

```
options ls=76 ps=62 ;
data;
input dpoint x1 x2 x3 y1 y2 y3;
array ys{3} y1-y3;
do i=1 to 3;
y = ys{i};
output;
end;
keep x1 x2 x3 y;
cards;
1 1 0 0 4.3 4.7 4.8
2 .5 .5 0 6.3 5.8 6.1
3 0 1 0 6.5 6.2 6.1
5 0 0 1 6.9 7 7.4
6 .5 0 .5 6.1 6.5 5.9
7 .34 .33 .33 6 5.8 6.4
8 .72 .14 .14 5.4 5.8 6.6
9 .14 .57 .29 5.7 5 5.6
10 .14 .29 .57 5.2 6.4 6.4
run;

proc glm;
model y=x1 x2 x3 x1*x2 x1*x3 x2*x3 / noint;
run;

proc glm;
model y=x1 x2 x1*x2 x1*x3 x2*x3;
estimate 'beta1' intercept 1 x1 1;
estimate 'beta2' intercept 1 x2 1;
estimate 'beta3' intercept 1;
run;
```

One graphical approach to understanding a mixture experiment analysis, besides contour plotting, is to look at response trace plots. These plot the estimated response along the line from a vertex, through the centroid, to the opposite edge. One can look at the response trace lines for many variables at once. Any that are nearly flat indicate inactive components.

For screening designs, Snee and Marquardt (1976) recommend the following design. For q components, take

- q pure components,
- q interior points, half-way between the vertices and the centroid,
- 1 centroid, and
- q endpoints—all permutations of $(0, 1/(q-1), \ldots, 1/(q-1))$,

giving $3q + 1$ points altogether. The results can be analyzed by looking at the response trace plots.

In many mixture situations, there will be constraints. In this case, the entire simplex cannot be used, and the feasible region will be some polytope. One approach is called the extreme vertex design. All of the vertices of the polytope are used, as well as the centroid of the region, and possibly centroids of the various edges, faces, and so on.

One generally better approach is to use a D-optimal approach. D-optimality means minimizing the determinant of the $X'X$ matrix. We will discuss it in depth next class. The XVERT algorithm of Snee and Marquardt (1974) is the basis for many computer implementations of this approach.

Another good approach is the so-called Distance-Based Design. In this case the algorithm picks points that are spread out uniformly in the feasible region.

A final important aspect of constrained components is the pseudocomponent approach. We define new components by

$$x_i' = \frac{x_i - a_i}{1 - \sum a_i},$$

where a_i is the lower bound of x_i. If the feasible region has only lower bounds, the result is a new, full-size, simplex. If upper bounds are present as well, the new region will at least be simpler than the old one.

If there are process variables (e.g., temperature, cooking time) involved besides mixture variables, there are two usual approaches. The first is to transform the q mixture variables into $q - 1$ independent variables and then proceed in the usual way. One typically does this by using ratios of components. That is $r_1 = x_1/x_3$, $r_2 = x_2/x_3$. The second approach is to directly model the mixture components. The idea is to do a mixture experiment at each point of a factorial design. In any case, the matter is tricky: there are many terms in the model, the variances and covariances of the coefficients will be large, and the interpretation of significant terms can be unclear.

Future research in mixture designs will need to include

- Blocking
- Process Variable Problem
- "Best Design Criteria"

- Other Model Forms
- Education
- Graphical Display

Reference: J. A. Cornell (1990), *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data,* 2nd edition, John Wiley & Sons, New York.

LECTURE 10

Robust Design and Taguchi Methods

1. Robust Design ala Taguchi

G. Taguchi, a Japanese engineer, had a big effect on quality control and experimental design in the 1980s and 1990s.

Let's begin with the sources of product variation. Some example of the random influences or "noise," that affect a product's characteristics are

1. Manufacturing
   - Operator
   - raw materials
   - machine settings
   - environmental
2. Environmental (the customer's environment)
   - temperature
   - humidity
   - dust
   - load
3. Product Deterioration (Aging)

Taguchi suggested that "quality" should be thought of, not as a product being inside or outside of specifications, but as the variation from the target. Variation from the target can be broken into two components, production variation, and bias. *Picture*

To quantify quality loss, write T for the target value and Y for the measured value. We want $E(Y) = T$. Write $L(Y)$ for the loss (in dollars, reputation, customer satisfaction, ...) for deviation of Y from T. A popular choice for the loss function is

$$L(Y) = k(Y - T)^2,$$

where k is some constant. If $E(Y)$ really is T, then $E(L(Y)) = k\sigma^2$, where $\sigma^2 = \text{Var}(Y)$.

If the product is off target, so that $E(Y) = T + d$, then $E(L(Y)) = k(\sigma^2 + d^2)$.

Now consider the product development stages at which countermeasures against various sources of variation can be built into the product.

Development Stages	Sources of Variation: Environmental Variables	Product Deterioration	Manufacturing Variations
Product Design	O	O	O
Process Design	X	X	O
Manufactoring	X	X	O

O—Countermeasures possible
X—Countermeasures impossible

We can think of quality control during manufacturing as on-line quality control, while quality control efforts during product design and process design are off-line quality control. There are potentially bigger payoffs from off-line quality control.

Consider the example from Bell Labs. In designing a power circuit which is to have a target output voltage of 115V.
Draw picture.
Voltage depends on transistor gain nonlinearly. The engineers set the transistor gain at 350, where the voltage response curve was fairly flat. Then then adjusted the resistor to return the voltage to the target of 115.

Taguchi recommended a two-step design process, robust design followed by tolerance design. Robust Design is a technique that reduces variation in a product by reducing the sensitivity of the design of the product to sources of variation rather than by controlling their sources. Tolerance Design is concerned with how much variation of the design and noise factors is permissible. It is a method for determining tolerances that minimizes the sum of product manufacturing and lifetime costs. The basic idea is to set tolerances around nominal settingss identified by parameter design, not by convention.

One must first identifify the control parameters. These are sometimes also called design parameters. They are the product design characteristics whose nominal settings can be specified by the product designers.

Next, one must state the problem and objectives. It works best to have a session that includes all the interested parties, and not to work in isolation.

Here are the operational steps for robust design.

1. State your problem and objective.
2. List responses, control parameters, and sources of noise.
3. Plan the experiment.
4. Run experiment and predict improved parameter settings.
5. Run confirmation experiment.

If the objective is not met, then it's back to step (2). Otherwise, you can adopt the improved design.

For the response variable, one begins by identifying important measures that are being targeted. One also wants a numerical representation of variability attributed to "noise." It is sometimes called a performance statistic. Two commonly used measures are

$$-\log(s^2),$$

where s^2 is the sample variance, and

$$\log(\text{mean}^2/s^2).$$

In general, the higher the performance measure the better.

Because of the need to estimate s^2, we are led to designs with two aspects. First one creates a design with design parameters, called by Taguichi the Inner Array. Then one creates the array of noise factors, called by Taguichi the Outer Array. Thus for each setting in the design matrix, one runs one run for each setting in the noise matrix. The measurements of the performance characteristic are then grouped to give the performance statistic for that run of the design matrix.

The Inner Array is generally a nearly-saturated array, usually a fractional factorial or Plackett-Burman design. The same kind of designs are used for for the Outer Array.

In the analysis, we'll look for factors that affect the targeted response only, those that affect variability only, those that affect both, and those that affect neither.

Let's consider an example from Bell Labs. The first step in silicon wafer fabrication is the growth of a smooth epitaxial layer onto a polished silicon wafer. The epitaxial layer is deposited on wafers while they are mounted on a rotating spindle called a susceptor. The problem: high drop-out rate caused by deviation in thickness, both between and within wafers, from the target value of 14.5 microns. Objective: reduce nonuniformity of epitaxial layer, and keep average thickness close to 14.5 microns. List responses: Epitaxial thickness, with a target value of 14.5 microns, a current average of 14.5, and a current std dev of 0.4. Control parameters: susceoptor rotation direction, arsenic flow rate, deposition time, nozzle position, and deposition temperature. List noise: uneven temperature in Bell Jar, nonuniform vapor concentration, nonuniform vapor composition, deviation in control parameter settings.

After settling on a 2^{5-2} factorial for the control variables, the experimenters placed four wafers with five sampling points each in the susceptor. In this wasy, they got a total of 20 measurements of epitaxial thickness, using the same plan for every run.

For each run, they calculated the mean, the variance, and the "robustness statistics," namely $-\log(\text{variance}/\text{mean}^2)$. The results were

Run	Rot Dir	As flow rate	dep temp	dep time	nozzle	mean	robustness stat
1	clock	55	1210	low	2	13.860	2.780
2	clock	55	1210	high	6	14.888	3.545
3	clock	59	1220	low	6	14.037	3.725
4	clock	59	1220	high	2	14.757	2.665
5	osci	55	1220	low	6	13.914	3.149
6	osci	55	1220	high	2	14.415	2.631
7	osci	59	1210	low	2	13.972	2.637
8	osci	59	1210	high	6	14.878	2.961

A larger robustness statistic corresponds to improved performance.

Draw pictures.

There are a number of issues that go into the choice of experimental design. These include

1. Number of levels of factors
2. Number of factors
3. Factor Interactions
4. Modeling versus Pick the Winner

The experimental designs commonly used include

1. Orthogonal Arrays
2. Plackett-Burman Designs
3. Fractional factorials
4. Response surface designs

Incidentally, the phrase Taguchi Methods is a trademark held by the American Supplier Institute. It encompasses all work due to Taguchi including quality engineering methods such as

1. Parameter Design
2. Tolerance Design
3. On-line quality control
4. The loss function
5. Signal-to-noise ratio

According to the ASI, if an experiment adheres to the following guidelines then it is a Taguchi experiment:

1. Best selection of quality characteristics (go/no go is not good).
2. Maximum possible number of control factors.
3. Comparison of existing conditions with predicted optimum
4. Use of signal-to-noise ratios.
5. Use of loss function.
6. Minimum interactions among control factors.
7. Control factors and noise factors separated.
8. Use of orthogonal arrays.

Strong Points of Taguchi Methods

1. Think about quality regarding closeness to target, not specification limits.
    - target
    - loss function
2. Transmission of error.
3. Analyze variation in addition to location.

Issues of Concern about Taguchi Methods and Taguchi Experiments

1. Blind use of orthogonal arrays.
2. No regard for interactions.
3. “Analysis” leads toward “pick-the-winner” rather than modeling.
4. Signal-to-noise ratio.

Summary

1. Examine both target and variability.
2. Find settings of factors to reduce sensitivity of product to fluctuations in noise parameters.

3. Use experimental design to determine factors that influence variability at target settings.
4. Attention needs to be paid to interactions.
5. Additional attention needs to be given to the interactive nature of experimentation.
6. Additional analysis should be placed on modeling system.

2. Improving Robust Design from a statistical point of view

2.1. Experimental Designs. Although the designs for the inner array and the outer array are economical individually, when they are crossed together they are not as economic.

For example, consider the crossed array when control factors A, B, C, in a 2^{3-1} crossed with another 2^{3-1} using D, E, and F in the outer array. The sixteen observations end up with one degree of freedom each for: A, B, C, D, E, F, AD, AE, AF, BD, BE, BF, CD, CE, CF. In other words, all the degrees of freedom go to main effects and noise $\times$ control interactions. These interactions are crucial, but interactions among the control variables may be just as crucial.

We would like a design to allow estimability of a reasonable model in both control and noise variables. To do that, we must first specify a reasonable model.

If we write x for the control variables and their settings, and z for the noise variables, then we might consider the response surface model

$$y(x, z) = \beta_0 + x'\beta + x'Bx + z'\gamma + x'\Delta z + \epsilon.$$

Notice that noise$\times$noise interactions are left out here.

We could use a crossed design for this model, but we could also use a "combined array," chosen specifically for this sort of model. The designs generally offer the concept of mixed resolution. For example, suppose there are three control factors A, B, C and three noise factors D, E, F. The usual design would be a crossed array

$$2^{3-1}_{III} \times 2^{3-1}_{I} II$$

with a total of 16 runs. We could view this as a single array with defining relations $I = ABC = DEF = ABCDEF$. A better alternative with the same number of runs would be a 2^{6-2} factorial with defining relations $I = ABCD = DEF = ABCDE$. This is resolution III for noise$\times$noise interactions and resolution IV for other interactions.

For a second order example, consider suppose we have 3 control variables and 2 noise variables. The crossed array might be a 1/3 fraction of a 3^3 and the noise array a 2^2, giving a total of 36 runs. A CCD with a 2^{5-1}, axial points in the control variables, and n_c center runs will total $22 + n_c$ runs.

2.2. Analysis. Taguchi's signal-to-noise approach, although easy to understand, is really sub-optimal. Modeling both the mean response and variance directly seems a much better idea. One can do this by:

1. Modeling the response as a function of control and noise variables, and then calculating the variance function from that, or

2. Take advantage of the crossed design, to calculate the variance at each control variable design point, and then simultaneously model the mean response and the variance. A classic paper by Bartlett and Kendall in 1946 suggests that using a log-linear model of the form

$$\log s_i^2 = x_i'\gamma + \epsilon,$$

will have approximately normal errors with constant variance.

Returning to the first point, from the given model the expected response, with the expectation taken over the distribution of the noise parameters, is

$$E(y(x, z)) = \beta_0 + x'\beta + x'Bx.$$

The variance is

$$\mathrm{Var}(y(x, z)) = \mathrm{Var}(z'\gamma + x'\Delta z) + \sigma^2 = \sigma_z^2(\gamma' + x'\Delta)(\gamma + \Delta' x) + \sigma^2).$$

Note that $(\gamma' + x'\Delta) = \partial y/\partial z$.

The text covers this point and many others quite thoroughly in Chapter 10.

LECTURE 11

Optimal Design

The issue of how to optimally design experiments has been around for a lont time, extending back to at least 1918 (Smith).

Before we start discussing the highlights of this topic we need to set up the problem. Note that we are going to set it up for the types of situations that we have enountered, although in fact the problem can be set up in more general ways.

We will define $X^{(n\times p)}$ as our design matrix, $\beta^{(p\times 1)}$ our vector of regression parameters, $y^{(n\times 1)}$ our vector of observations, and $\epsilon^{(n\times 1)}$ our error vector giving

$$y = X\beta + \epsilon.$$

We assume ϵ will be iid with mean zero and $\text{Cov}(\epsilon) = \sigma^2 I$. As before, we have

$$\hat{\beta} = (X'X)^{-1}X'y$$
$$\text{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$$
$$\hat{y}_x = x\hat{\beta}$$
$$\text{Var}(\hat{y}_x) = \sigma^2 x(X'X)^{-1}x'.$$

The design problem consists of selecting row vectors $x^{(1\times p)}$, $i = 1, 2, \ldots, n$ from the design space $\mathcal{X}$ such that the design defined by these n vectors is, in some defined sense, optimal. We are assuming that n is fixed. By and large, solutions to this problem consist of developing some sensible criterion based on the above model and using it to obtain optimal designs.

One of the first to state a criterion and obtain optimal designs for regression problems was Smith(1918). The criterion she proposed was: minimize the maximum variance of any predicted value (obtained by using the regression function) over the experimental space. I.e.,

$$\min_{x_i, i=1,\ldots,n} \max_{x\in\mathcal{X}} \text{Var}(\hat{y}_x).$$

This criterion was later called global, or G-optimality by Kiefer and Wolfowitz (1959).

A second criterion, proposed by Wald (1943), puts the emphasis on the quality of the parameter estimates. The criterion is to maximize the determinant of $X'X$. That is

$$\max_{x_i, i=1,\ldots,n} |X'X|.$$

This was called D-optimality by Kiefer and Wolfowitz (1959). Comment on confidence ellipsoids: the determinant is the product of the eigenvalues, which is inversely proportional to the product of the axes of the confidence ellipsoid around $\hat{\beta}$, so maximizing $|X'X|$ is equivalent to minimizing the volume of the confidence ellipsoid.

In their General Equivalence Theorem, the equivalence of D and G optimality was established under certain conditions (to be discussed shortly).

While these criteria are the ones which have received the most attention in the literature, others have also been used. For example, so called A-optimality,

$$\min_{x_i, i=1,\ldots,n} \text{trace}(X'X^{-1}),$$

minimizes the average variance of the parameter estimates (Chernoff, 1953). Another criterion, E-optimality, finds the design which maximizes the minimum eigenvalue of $X'X$ (Ehrenfeld, 1955). A conceptually attractive criterion is called V-optimality (sometimes IV-optimality or Q-optimality). Here the criterion is to minimize the integrated prediction variance over the region of interest.

Of all these designs, only D-optimality is invariant under reparametrization.

Since D and G optimality are the criteria receiving the most attention in the applied literature, we will have a more detailed discussion of these criteria. Before doing so, we need to make a distinction between what is called the exact theory and the approximate theory. Suppose you had a problem involving maximizing a function over the integers. Standard calculus techniques don't apply. A common technique would be to extend the function definition to the real numbers, use calculus to find the number where the maximum occurs, and then argue that the maximum over the integers will occur at an adjacent integer. The analogous design problem distinguishes the exact theory (like the integers) from the easier approximate theory (like the reals).

All of the design criteria just discussed have the property that

$$\phi(aX'X) = \text{positive constant} \times \phi(X'X),$$

so a design that maximizes $\phi(aX'X)$ also maximizes $\phi(X'X)$. Suppose that we have an n-point design with n_i observations at x_i, so that $\sum n_i = n$. This. or any, design can be viewed as a measure ξ on the design space $\mathcal{X}$. Let ξ be a probability measure on $\mathcal{X}$ such that

- $\xi(x_i) = 0$ if there are to be no observations at x_i, and
- $\xi(x_i) = n_i/n$ if there are to be $n_i > 0$ observations at x_i.

For a discrete n-point design ξ takes on values which are multiples of $1/n$, and defines an exact design on $\mathcal{X}$.

If we remove the restriction that ξ be a multiple of $1/n$, we can extend this idea to a design measure which satisfies

$$\xi(x) \geq 0, \qquad x \in \mathcal{X}$$

$$\int_{\mathcal{X}} \xi(dx) = 1.$$

Now let

$$m_{ij}(\xi) = \int_{\mathcal{X}} x_i x_j \xi(dx), \qquad \text{for all } i,\ j = 1, \ldots, p,$$

where $m_{ij}(\xi)$ is the ij element of the matrix $M(\xi)$. Note, for an exact design (this is called the moment matrix),

$$M(\xi) = \frac{1}{n} X'X.$$

Similarly, a normalized generalization relating to $\text{Var}(\hat{y}_x)$ is

$$d(x, \xi) = x(M(\xi))^{-1}x',$$

again for an exact design

$$d = d(x, \xi) = nx(X'X)^{-1}x'.$$

Using this notation we have the following definitions,

- ξ^* is D-optimal if and only if $M(\xi^*)$ is nonsingular and

$$\max_{\xi} |M(\xi)| = |M(\xi^*)|.$$

- ξ^* is G-optimal if and only if

$$\min_{\xi} \max_{x \in \mathcal{X}} d(x, \xi) = \max_{x \in \mathcal{X}} d(x, \xi^*).$$

It turns out that a sufficient condition for ξ^* to satisfy the G-optimality criterion is

$$\max_{x \in \mathcal{X}} d(x, \xi^*) = p,$$

where p is the dimension of $M(\xi^*)$, or equivalently, the number of parameters in the model. To see that p is a lower bound for $\max d(x, \xi)$, consider

$$\begin{aligned} p &= \text{trace} M M^{-1} \\ &= \frac{1}{n} \text{trace} X'X M^{-1} \\ &= \frac{1}{n} \sum_i \text{trace}(x_i^t x_i) M^{-1} \\ &= \sum_i \text{trace}(x_i M^{-1} x_i^t) \\ &\leq \max_{x \in \mathcal{X}} x M^{-1} x^t \end{aligned}$$

How would one show that a specific design is the best there is? The key is to look at the derivative. For a given design M and a given optimality criterion ϕ to maximize, like log det, the Fréchet derivative is defined as

$$F_\phi(M, x^t x) = \lim_{\epsilon \to 0^+} \frac{1}{\epsilon} \left[\phi\left((1-\epsilon)M + \epsilon x^t x\right) - \phi(M) \right].$$

For example, let's consider D-optimality. We want to maximize the determinant of M. This is equivalent to maximizing the log of the determinant of M. The function

$\phi = \log\det$ has the advantage that it is convex on the space of information matrices M, so that a local maximum will in fact be a global maximum. At any rate

$$\begin{aligned}
F_\phi(M, x^t x) &= \lim_{\epsilon\to 0^+} \frac{1}{\epsilon}\left[\phi\left((1-\epsilon)M + \epsilon x^t x\right) - \phi(M)\right] \\
&= \lim_{\epsilon\to 0^+} \frac{1}{\epsilon}\left[\log\det\left((1-\epsilon)M + \epsilon x^t x\right) - \log\det(M)\right] \\
&= \lim_{\epsilon\to 0^+} \frac{1}{\epsilon}\left[\log\frac{\det\left((1-\epsilon)M + \epsilon x^t x\right)}{\det M}\right] \\
&= \lim_{\epsilon\to 0^+} \frac{1}{\epsilon}\left[\log\det\left((1-\epsilon)I + \epsilon x^t x M^{-1}\right)\right] \\
&= \lim_{\epsilon\to 0^+} \frac{1}{\epsilon}\left[\log(1-\epsilon)^p \det\left(I + \frac{\epsilon}{(1-\epsilon)} x^t x M^{-1}\right)\right] \\
&= \lim_{\epsilon\to 0^+} \frac{1}{\epsilon}\left[\log(1-\epsilon)^p \left(1 + \frac{\epsilon}{(1-\epsilon)}\mathrm{trace}(x^t x M^{-1}) + \mathcal{O}(\epsilon^2)\right)\right] \\
&= \lim_{\epsilon\to 0^+} \frac{1}{\epsilon}\left[\log(1-\epsilon)^p + \log\left(1 + \frac{\epsilon}{(1-\epsilon)}\mathrm{trace}(x^t x M^{-1}) + \mathcal{O}(\epsilon^2)\right)\right] \\
&= \lim_{\epsilon\to 0^+} \frac{1}{\epsilon}\left[p\log(1-\epsilon) + \frac{\epsilon}{(1-\epsilon)}\mathrm{trace}(x^t x M^{-1}) + \mathcal{O}(\epsilon^2)\right] \\
&= x^t M^{-1} x - p.
\end{aligned}$$

(Remember that $\log(1+t) = t + \mathcal{O}(t^2)$.)

Why have we only considered the derivative of M in the direction of a matrix of the form $x^t x$ and not something more general? As a consequence of Carathéodory's Theorem, every element of the design space can be expressed as a convex combination of no more than $p(p+1)/2 + 1$ elements of the form $x^t x$.

Finally, the equivalence of D- and G-optimality is established in the General Equivalence Theorem of Kiefer and Wolfowitz. The General Equivalence Theorem says: If ϕ is concave on $\mathcal{M}$, the space of design information matrices, and differentiable at $M(\xi^*)$, then the following are equivalent

1. The measure ξ^* is ϕ-optimal
2. The Fréchet derivative $F_\phi(M(\xi^*), x^t x) \le 0$ for all $x \in \mathcal{X}$.
3. The following equality holds

$$\max_{x\in\mathcal{X}} F_\phi(M(\xi^*), x^t x) = \min_{\xi} \max_{x\in\mathcal{X}} F_\phi(M(\xi), x^t x).$$

This last is what gives the equality of D-and G-optimality.

The implication of this result is that we can use the sufficient condition for G-optimality to verify whether or not a specific design is D-optimal. That is, if

$$\max_{x\in\mathcal{X}} d(x, \xi^*) = p,$$

where as before p is the number of parameters in the the model, including the intercept, then the design is D-optimal.

Note that D-optimality is essentially a parameter estimation criterion, whereas G-optimality is a response estimation criterion. The Equivalence Theorem says that these two design criteria are identical when the design is expressed as a measure on $\mathcal{X}$.

Note that when dealing with exact designs the equivalence of the two criteria does not hold.

In practice, the design problem consists of selecting an exact design to define the experimental runs. The design measure that we just discussed gives us an approximate design.

Measure designs are of interest primarily because the D-optimal measure design provides the reference against which exact designs can be evaluated, and also because the points in an optimal exact design will often correspond to the points of support (points of positive measure) of the D-optimal measure design.

For the practical problem we will let ξ be an n-point design. The moment matrix

$$M(\xi) = \frac{1}{n} X'X$$

and

$$|M(\xi)| = \frac{1}{n^p} |X'X|.$$

Again a normalized measure of the variance of the prediction at x is

$$d = x(M(\xi))^{-1}x' = nx(X'X)^{-1}x'.$$

Note, with our usual assumption,

$$\text{Var}(\hat{y}_x) = \frac{1}{n}\sigma^2 d.$$

The values of $|M(\xi)|$ and d give an indication of the information per point for a design, so designs having differing numbers of points can be compared. Then designs can be compared based on their D- and G-efficiencies.

For a given design, call it ξ', define the D-efficiency to be

$$\left[\frac{|M(\xi')|}{\max_\xi |M(\xi)|}\right]^{\frac{1}{p}}.$$

Similarly, G-efficiency is defined as

$$\frac{p}{\max_{x\in\mathcal{X}} d(x,\xi)} = \frac{p}{nd}.$$

One should note that for finite designs, especially small ones, the efficiency is likely to be quite a bit less than 1.

Now let's briefly consider a few examples of D-optimal designs. First note in the above presentation, that two pieces of information must be supplied prior to obtaining an optimal design. These are:

1. model to be used
2. Number of data points

We could also, as Kiefer did, talk about the location of points and proportion of points at that location, rather than the number of points, but that is not the usual practice in designing experiments.

Some examples:

1. Consider the model

$$E(y) = \beta_0 + \sum \beta_i x_i + \sum \sum_{i \le j} \beta_{ij} x_i x_j + \beta_{123} x_1 x_2 x_3,$$

with $i, j, k = 1, 2, 3$, so $p = 8$. Consider all possible designs with eight points (so $n = 8$) with the restriction

$$|x_i| \le 1.$$

It can be shown that, for fixed diagonal terms, $|X'X|$ is largest when all off-diagonal terms are zero. This can be achieved by adopting a 2^3 design with points $(\pm a, \pm a, \pm a)$ with $|a| \le 1$. For this design it can be shown that

$$|X'X| = n^p a^{2(3+3\cdot 2+3)} = 8^8 a^{24}.$$

Clearly this is maximized if $|a| = 1$. That, in turn, implies that the optimal design is 2^3 with $a = \pm 1$. Also notice that $d = nx(X'X)^{-1}x' = 8 = p$.

2.

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3.$$

Here $p = 4, n = 4$. Using the same type of argument as in example 1, the optimum design is 2^{3-1} with $I = 123$.

3.

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \beta_{112} x_1^2 x_2 + \beta_{122} x_1 x_2^2 + \beta_{1122} x_1^2 x_2^2.$$

Here $p = 9$ and $n = 9$. The optimal design is the 3^2 factorial with levels (-1,0,1).

4.

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2,$$

the full second-order model in two variables. Here $p = 6$ and suppose $n = 6$. The optimal design is

x_1	x_2
-1	-1
1	-1
-1	1
$-\alpha$	$-\alpha$
1	3α
3α	1

where $\alpha = 0.1315$. These designs may be rotated by 90^o or a multiple of that.

5. Using the same model with $n = 9$ the optimal design is 3^2.
6. Using a full second-order model and the number of points for a central composite for $p \ge 2$, the D-optimal design places 2^p points at ± 1 and the star points are placed on the face of the hypercube.

Note that in the last two designs, the design space was a hypercube, not a hypersphere.

A theory, no matter how beautiful, won't be used much in practice if it's not practical. For optimal designs, this comes down to a question of having algorithms to find optimal designs. These were first worked out for D-optimality in the early 1970s. The basic idea is quite simple. Suppose you have a design with information matrix M_n. Find x_{n+1} to maximize $F(M_n x x^t)$. Unless M_n is optimal, this value will be positive, and you can increase the design criterion measure by moving from M_n in the direction of $x_{n+1}x_{n+1}^t$. Thus for some α_{n+1} we define a new measure by

$$\eta_{n+1} = (1 - \alpha_{n+1})\eta_n + \alpha_{n+1}\eta_{n+1},$$

with corresponding information matrix

$$M_{n+1} = (1 - \alpha_{n+1})M_n + \alpha_{n+1}x_{n+1}x_{n+1}^t.$$

The step-length α_n can either be chosen to maximize the criterion along that ray, as was done by Federov (1972), or as a sequence converging to zero but with divergent partial sums, as was done by Wynn (1970).

In particular, for D-optimality,

$$F(M, xx^t) = x^t M^{-1} x - p.$$

Writing $d_n = x^t M_n^{-1} x$ and $\bar{d}_n = \max x^t M_n^{-1} x = x_{n+1}^t M_n^{-1} x_{n+1}$, the optimal step-size turns out to be

$$\alpha_{n+1} = \frac{\bar{d}_n - p}{p(\bar{d}_n - 1)}.$$

Further comments on algorithms.

SAS has `proc optex` to generate optimal designs. Typically one gives it a set of candidate points and a model. For example, suppose you've done a resolution IV fractional factorial design on seven factors. You'd like to augment the design so that you can estimate all two-factor interactions. First generate a set of candidate points:

```
proc factex;
factors x1-x7;
output out=can;
run;
```

and now the resolution IV design:

```
proc factex;
factors x1-x7;
model resolution=4;
size design=min;
output out=aug;
run;
```

Finally, find the augmented design totalling 30 points, according to the D-optimality criterion:

```
proc optex data=can
model x1|x2|x3|x4|x5|x6|x7@2;
generate n=30 augment=aug;
output out=design;
run;
```

You will note that the criteria, and therefore the optimal designs that are developed, are model dependent. A question that arises is how well do these designs perform if the model is incorrect. We discussed that earlier in the lecture about bias and variance.

In fact, when using a computer to find an optimal design, one should always keep in mind that

1. The designs are model dependent, and may not be particularly good for other models.
2. Exact designs for D-optimality do not address prediction variance; the equivalence of D- and G-optimality does not hold for exact designs.
3. D-optimal designs do not allow for many center runs.

In general, for a situation that can be handled with a Central Composite Design or Box-Behnken, those are better designs in an overall sense. Nonetheless, computer-aided optimal designs can be invaluable for

1. Mixture designs
2. Constrained designs
3. Trying to salvage a "botched" experiment

References Although the literature on optimality is vast (over 600 articles), I'll stick to three books, in increasing order of difficulty:

A. C. Atkinson and A. N. Donev, *Optimum Experimental Designs,* Oxford University Press, 1992.

S. D. Silvey, *Optimal Design*, Chapman and Hall, 1980. (Just 86 pages but unfortunately out of print.)

F. Pukelsheim, *Optimal Design of Experiments*, Chapman and Hall, 1995.

LECTURE 12

Computer Experiments

The advent of computers has had a profound effect on the design of experiments. Most obviously, it is enormously easier to analyze an experiment. Subjects like the theory of optimal designs (e.g. D-optimal designs) didn't spread beyond journal articles until computer algorithms became available. In the last twenty years, the arrow has also gone the other way as statisticians have looked at experiments done on computers, as opposed to the physical world.

Physical phenomena in science and engineering are widely simulated on computers. Computers are used to model the flow of air over aircraft wings, behaviour of metal structures under stress, nuclear reactor safety, and many other situations. Probably the most important is the simulation of semiconductors. A process simulator simulates taking unprocessed silicon through the steps of oxidation, etching, and ion injection. A device simulator begins with the description of a semiconductor device and simulates current flowing through it under varying conditions. A circuit simulator begins with a list of such devices and their arrangement and simulates the properties of the circuit as a whole.

In all these simulations, some variables are specified by the user. Given those variable settings, the results are fixed. That is, there is no random variation in the response. The number of variables can vary from under a dozen to several hundred. The number of response variables is generally greater than one, and the time for one run of the simulation can vary from milliseconds on a PC to several hours on a supercomputer.

There are a variety of possible goals in computer experiments.

1. Finding a good value of X according to some criterion of Y,
2. Finding a simple approximation $\hat{f}$ that is accurate enough in the region of interest, but is far easier to calculate or understand,
3. Estimating the size of the error $\hat{f}(X_0) - f(X_0)$,
4. Sensitivity of Y with respect to changes in X,
5. Finding which components of X are most important for each component of the response Y,
6. Visualizing f,
7. Uncovering bugs in the simulation.

As to (1), standard optimization methods (quasi-Newton or conjugate gradients) work well if one needs the precise optimum and has a good starting value, but computer experiments can be useful is searching for a good starting place, or if

one is looking for widely separated regions of the design space that might all have good response values.

For convenience we will assume that every design variable is continuous and runs from 0 to 1.

There are two approaches to the subject, Bayesian and frequentist. The Bayesian approach is elegant and well developed. The disadvantage is computation: with n runs, one must solve n equations in n unknowns. Thus the effort grows as n^3. Thus if one needs an hour for the computer simulations and one minute to solve the equations, and then decides on increasing n by a factor of 24, one will need one day for the simulations and 9.6 days for the equations.

1. The Bayesian Approach

The usual approach is the Kriging model. Here we consider

$$Y(x) = \sum_{j=1}^{k} \beta_j h_j(x) + Z(x),$$

where the h_j are known fixed functions, the β_j are unknown coefficients to be estimated, and Z is a random function with 0 expectation and covariance dependent on the separation of the two design points:

$$\mathrm{Cov}(Z(x_i), Z(x_j)) = \sigma^2 R(x_j - x_i).$$

Typical choices for R include $R(d) = \exp(-\theta|d|)$, $R(d) = \exp(-\theta|D|^q)$ (called Gaussian when $q = 2$, and a cubic in $|d|$. Obviously the choice of R has a big effect on the conclusions.

At any rate, given all responses y_D over a design, we can consider the predictor

$$\hat{y}(x_0) = \lambda'(x_0) y_D.$$

What predictor of this form does best? It is not to difficult to find the Best Linear Unbiased Predictor, but I will omit it here. It has two terms, a generalized least squares predictor for the point x_0, and a term that pulls the response surface through the observed points.

As for a choice of experimental design, we are in the unusual position of having no random error, only bias, to consider. There are four standard design criteria:

1. Entropy, namely maximize the expected information, or in the Gaussian case, maximize the determinant of the variance of Y_D
2. Mean squared error
3. Minimax, minimize the maximum distance to a design point
4. Maximin, maximize the minimum distance between any two design points.

Entropy designs tend to spread the points out in the design space and the favor the boundaries over the interior. Mean squared designs favor the interior and tend to clump when projected onto lower dimensions.

2. The frequentist approach

Here we model Y as

$$Y = f(X) \approx Z(X)\beta.,$$

where $Z(X)$ is a row vector of predictor functions and β is a vector of parameters. Z might be made out of low-order polynomials, trig polynomials, wavelets, or other functions suitable to the application. The usual way criterion for estimating β is to minimize the mean squared error of the approximation with respect to some distribution on the design space.

If we pick the functions to be orthogonal and assume a uniform distribution on the design space, then

$$\hat{\beta} = \frac{1}{n} \sum Z(x_i)' f(x_i).$$

Here computations grow at a rate of n^2, so in the earlier example, when the simulations take a day, the estimate of $\hat{\beta}$ takes 9.6 hours. The variance of $\hat{\beta}$ is

$$\frac{1}{n} \operatorname{Var}(Z(X)'Y(X)).$$

A good design will allow good quality numerical integration and allow good estimation of the sampling variance. Iid sampling, for example, makes variance estimation easier but reduces the accuracy of the numerical integration. There is also an issue of getting points into corners of the design space so that important features are not missed.

One approach is to use a regular grid, say k values for each component of X. This grows rather quickly, and the grid is poor for variance estimation. One could instead pick "good lattice points," perhaps with some random perturbations These work quite well for smooth periodic functions.

Latin hypercubes are the predominant frequentist design. The design points are determined by

$$X_i^j = \frac{\pi^j(i) - U_j^i}{n},$$

where π^j are uniform random permutations of the integers 1 through n, and U_i^j are iid $U[0, 1]$. Sometimes U_j^i is just set to 0.5. At any rate, the sample points are stratified on each of the input variable axes. This stratification reduces the variance of the estimated integrals. A related approach is to use

$$X_i^j = \frac{\pi^j(A_i^j) - U_j^i}{n},$$

where A_i^j is an element of an orthogonal array.

A good reference (from which this lecture was drawn) is

J. B. Koehler and A. B. Owen, "Computer Experiments," in *Handbook of Statistics, Vol 13*, S. Ghosh and C. R. Rao, eds., Elsevier Science B. V., 1996.

LECTURE 13

Gene Expression Arrays

1. Introduction

Microarrays are used by biologists to study what genes are expressed in certain tissue. The array consists of samples of DNA of known sequence spotted onto a glass slide. Purified mRNA from the tissue under study is converted into cDNA and tagged with red or green flourescent dye, which is then washed across the slide. Strands of cDNA that match the spots binds, while other cDNA washes off. The red and green intensities from a spot indicate the relative abundance of the mRNA in the original tissue samples; the more mRNA, the more gene expression.

Microarrays are a hot new tool in biology. The analysis of the data from them has aroused a lot of interest from statisticians. Experimental design using microarrays is also an interesting problem, but one which has attracted considerably less interest, with the notable exception of Gary Churchill at Jackson Labs.

There are a variety of factors of interest in a microarray experiment: varieties (V), whether tumor versus non-tumor tissue, the tissue over time, the tissue exposed to a drug, and so on; the genes (G), corresponding to the spots on the array; the dyes (D), meaning red versus green, and the arrays (A), since quality control of microarrays is far from perfect.

There are a number of two-way interactions of interest. First, the dyes are bound to the cDNA in separate runs. Differences could result in a DV interaction. The same gene on different arrays can have different amounts of cDNA, giving an AG interaction. Occasionally one sees specific genes bind better to one dye than the other, yielding a DG effect. Finally, the interaction of biological interest is the variety by gene, VG, interaction.

2. Designs

The most common design is the so-called "reference" design. One dye is the reference variety, the other dye is used for the varieties of interest. Observe that V is confounded with G, meaning VG is confounded with DG. Note also that the reference variety parameters are the best estimated.

With v varieties in addition to the reference, and n genes, one will have $2vn$ observations. The mean, A, V, and G account for $2v + (n-1)$ degrees of freedom. VG has $v(n-1)$ degrees of freedom. The final $(v-1)(n-1)$ degrees of freedom go to the AG interaction (or if AG is ignored, they go to error).

An alternative is a called a "loop" design. With the same number of arrays as the reference design, the loop gives twice as much information on the non-reference varieties, and varieties are balanced with respec to dyes. That is, VG is unconfounded with DG.

With a loop design, after all factor main effects plus VG and AG interactions, there are $n - 1$ degrees of freedom left for error.

3. Model

A classic ANOVA model for microarray data might be

$$y_{ijkg} = \mu + A_i + D_j + V_k + G_g + (VG)_{kg} + (AG)_{ig} + \epsilon_{ijkg}.$$

(Adding a $(DG)_{jg}$ term is also possible.) There's really only one type of contrast of interest: $(VG)_{k_1g} - (VG)_{k_2g}$. That suggests using an optimality criterion of minimizing the average variance of $\sum \mathrm{Var}\,\hat{(}VG)_{k_1g} - \hat{(}VG)_{k_2g}$

Reference: M. K. Kerr and G. A. Churchill, "Experimental Design for Gene Expression Microarrays," *Biostatistics* **2**, 2001, 183–201.

LECTURE 14

Miscellany

1. Errors in design parameters

Even in the lab, in can be difficult to precisely control the levels of the various design parameters. What is the impact of errors of control of design levels? The answer depends on the structure of the errror. Probably the most common instance is when the experimenter decides on a level which is ostensibly used. That exact level is used for the analysis. The actual level from the process's point of view is unknown (Myers and Montgomery, on whom this section is heavily drawn, call this the "fixed design level" case.

The deviation between the actual and planned level is a random variable. For the u-th run of the ith design variable we have

$$w_{ju} = x_{ju} + e_{ju},$$

where w_{ju} is the actual level, x_{ju} is the planned level, and e_{ju} is the error in control, not to be confused with ϵ, the usual error in y. The usual assumption is that

$$E(e_{ju}) = 0$$
$$\text{Var}(e_{ju}) = \sigma_j^2$$

Let us first consider a first order model. The model is

$$\begin{aligned} y &= \beta_0 + \beta_1 w_1 + \beta_2 w_2 + \cdots + \beta_k w_k + \epsilon \\ &= \beta_0 + \beta_1(x_1 + e_1) + \beta_2(x_2 + e_2) + \cdots + \beta_k(x_k + e_k) + \epsilon \\ &= \beta_0 + \sum \beta_j x_j + (\epsilon + \sum \beta_j \theta_j) \\ &= \beta_0 + \sum \beta_j x_j + \epsilon^*. \end{aligned}$$

Thus we can think of just having a larger error term. If we use this last equation, and estimate parameters using least squares, we'll have the following

1. The regression coefficients are unbiased.
2. The covariance matrix of $\hat{b} = (X'X)^{-1}X'y$ is $(X'X)^{-1}\sigma^{*2}$, where $\sigma^{*2} = \sigma^2 + \sum \beta_j^2 \sigma_j^2$ (assuming the θ_j are independent.
3. Standard response surface methods, such as screening, steepest ascent, and modeling, remain valid, although if the β_j are large they may be less effective.

Now consider a second order model. We have

$$\begin{aligned}
y &= \beta_0 + \sum \beta_j w_j + \sum \beta_{jj} w_j^2 + \sum\sum \beta_{ij} w_i w_j + \epsilon \\
&= \beta_0 + \sum \beta_j (x_j + e_j) + \sum \beta_{jj} (x_j + e_j)^2 + \sum\sum \beta_{ij} (x_i + e_i)(x_j + e_j) + \epsilon \\
&= \beta_0 + \sum \beta_j x_j + \sum \beta_{jj} x_j^2 + \sum\sum \beta_{ij} x_i x_j \\
&\quad + \left(\epsilon + \sum \beta_j e_j + \sum \beta_{jj} e_j (2x_j + e_j) + \sum\sum \beta_{ij} (e_i x_j + x_i e_j + e_i e_j)\right) \\
&= \beta_0 + \sum \beta_j x_j + \sum \beta_{jj} x_j^2 + \sum\sum \beta_{ij} x_i x_j + \epsilon^*.
\end{aligned}$$

Clearly ϵ^* is more complicated than in the linear model. First,

$$E(\epsilon^*) = \sum \beta_{jj} \sigma^2.$$

Note that this doesn't depend on x_j. Thus the bias affects only the β_0 term.

Before looking at the variance, we rewrite ϵ^* as

$$\begin{aligned}
\epsilon^* = \epsilon + \sum\sum\nolimits_{i \neq j} e_j \beta_{ij} x_j + \sum\sum \beta_{ij} e_i e_j \\
\sum e_j \beta_j + 2 \sum \beta_{jj} e_j x_j + \sum \beta_{jj} e_j^2.
\end{aligned}$$

Clearly,

1. The error variance is inflated, just as in the linear case
2. The error variance is no longer homogeneous, but depends on the design levels.

For further reference, see the article by Box (1963) in *Technometrics*, or for a more general discussion of measurement error, the book by Fuller or Carrol and Rupert.

2. Split Plots

Our designs and analyses have been based on completely randomized run orders. Sometimes, or even often, doing so is hard or impossible. For example, if one of the factors is oven temperature, most experimenters would want to fix the oven temperature, vary all the other factors, bake a batch, change the temperature, and so on. Indeed, Taguchi designs are often run with fixed control settings and varying environmental settings.

Designs with such hard to change factors can be run as "split plots." (Blocks might be a better term, but we already used that. The term "split plot" comes from agriculture.) A split plot design traditionally has a factorial structure, but with two different experimental units called whole plots and subplots. These two units require different randomization and analysis. The difficult to change factor has levels randomly assigned to whole plots. For each whole plot, the easier to change factor levels are then randomly assigned to subplots. Suppose oven temperature (T) is difficult to change and concentration (C) is easy to change, and that both have two levels. Think of each replication of the design as looking like

Whole Plot 1	Whole Plot 2
T, C	T,C
-,-	+,-
-,+	+,+

Clearly temperature is confounded with blocks. If this design were replicated r times, then the ANOVA table is

Source	df
Replications	r-1
Temperature	1
Temperature×replication	r-1 (whole plot error)
Concentration	1
Concentration×Temperature	1
Concentration×replications	r-1
Concentration×temperature×replication	r-1
Total	4r-1

The last two components in the table are together known as the subplot error.

Notice that

1. There are more degrees of freedom for estimating the subplot effects than for the whole plot effects, so any tests on subplot effects will be more efficient.
2. We could think of the subplot levels as a secondary replicated randomized design

Notice that if we ignored the randomization scheme we would have the following ANOVA table

Source	df
Temperature	1
Concentration	1
Temperature×Concentration	1
Error	4(r-1)
hline Total	4r -1

Now consider a general situation with hard to change variables $z_1, z_2, \ldots$ and more standard variables $x_1, x_2, \ldots$, and the randomization has been done with a split plot structure. This requires all combinations of the whole plot factors to be crossed with the same combination of the subplot factors. An example would be a complete factorial. Say there are a unique combinations of the z's, and b combinations of the x's. The z-combinations are randomly assigned. Next, for each z-combination, the x-combinations are randomly assigned. For example, if we have a single z-factor, denoted with A, and three x-factors, denoted with B, C, D, the split plot structure would look like

(1)	a
b	ab
c	ac
bc	abc
d	ad
bd	abd
cd	acd
bcd	abcd

The general model is then

$$y_{ij} = \beta_0 + \gamma' z_j + \beta' x_{ij} + z_i' B_z z_j + x_{ij}' B_x x_{ij} + z_i' \Delta x_{ij} + \delta_i + \epsilon_{ij},$$

where $\delta_i \sim N(0, \sigma_\delta^2)$ and $\epsilon_{ij} \sim N(0, \sigma^2)$.

Unfortunately, replication is a luxury in industrial experiments, so split plot designs are problematic. For example, two observations from the same whole plot share a common error component, and so have covariance σ_δ^2. In fact, the block of the covariance matrix corresponding to one whole plot has $\sigma^2 + \sigma_\delta^2$ on the diagonal and σ_δ^2 on the off-diagonal, while the covariance across whole plots is zero. Parameter estimation in this situation depends on using "generalized least squares." This in turn requires estimation of σ^2 and σ_δ^2. For similar reasons, half- or full-normal plots for screening should be done separately for whole- and sub-plot effects. On the other hand, in Taguchi style experiments, one usually just wants the variance at each control-factor design point, so these issues don't really matter.

Box and Jones (1992), *J. Appl Stat*, Lucas and Ju (1992) it ASQC Quality Congress Trans., and Letsinger et al. (1996) *J. Qual Tech* are the only discussions of these issues that I know of.

3. Qualitative factors

Intro

Let us start with two levels of the qualitative factor. Qualitative factors are usually modeled with dummy variables. While one could use 0 and 1 for the factor levels, using -1 and 1 is consistent with our prior usage. As long as all qualitative factors have two levels, for first order designs it doesn't matter how many qualitative factors there are: fractional factorial designs work just dandy.

As an example, consider an extraction process in which the factors are time, temperature, and type of solvent. Solvents A and B are used, coded -1 and 1, respectively. The fitted first order model is

$$\hat{y} = 17.5 + 4.7x_1 + 10x_2 - 3z,$$

where x_1 is the variable for time, x_2 for temperature, and z for solvent. One way to interpret this model is to say that for solvent A, the model is

$$\hat{y} = 20.5 + 4.7x_1 + 10x_2,$$

and for solvent B the model is

$$\hat{y} = 14.5 + 4.7x_1 + 10x_2.$$

For multiple levels of a qualitative factor, we could take the approach discussed in one of the talks last week, or simply use the dummy variables. For example, is solvent had three levels, we could code this with two (0,1) variables z_1 and z_2, so that one level would be (0,0), the second (1,0), and the third (1,1). For four levels, we could use 3 variables and code (0,0,0) for the first level, (1,0,0) for the second, (0,1,0) for the third, and (0,0,1) for the last.

If there is no interaction between qualitative and quantitative factors, the qualitative factors can be thought of as only changing the model intercept. Otherwise, the situation is more complicated. Here's a first order plus interactions model of the preceeding example

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 + cz + d_1x_1z + d_2x_2z.$$

That translates to saying that for solvent A, the model is

$$\hat{y} = (b_0 - c) + (b_1 - d_1)x_1 + (b_2 - d_2)x_2 + b_{12}x_1x_2$$

and for solvent B the model is

$$\hat{y} = (b_0 + c) + (b_1 + d_1)x_1 + (b_2 + d_2)x_2 + b_{12}x_1x_2.$$

In general, for first-order designs with qualitative factors, one wants

1. the ability to estimate the parameters of the quantitive factors
2. to estimate the interaction between qualitative and quantitative terms
3. to be able to test for simpler models.

Consider the model

$$y = \beta_0 + beta_1x_1 + \beta_2x_2 + \gamma_1z + \epsilon.$$

A 2^3 factorial, with or without center points, would accomplish everything desired here. What if you only had 5 runs? Consider the following 5 designs

x_1	x_2	z	x_1	x_2	z	x_1	x_2	z	x_1	x_2	z	x_1	x_2	z
-1	-1	-1	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	-1	1
1	-1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	1
-1	1	-1	-1	1	-1	-1	1	-1	-1	1	1	-1	1	-1
1	1	-1	1	1	1	1	1	1	1	1	-1	1	1	-1
0	0	1	0	0	-1	0	0	-1	0	0	1	0	0	-1

All of these designs allow estimation of β_0 and β_1 and are at both levels of the qualitative factor. Furthermore, all allow estimation of $\beta_{12}x_1x_2$. The third and fifth designs also allow δ_1x_1z or δ_2x_2z to be estimated.

In general, two-level factorials and fractions with resolution V or better are best. With resolution V designs, all two-level interactions are estimable. Another approach is to use lower resolution designs that are high enough resolution in the qualitative factors. For example, with 5 quantative factors and 1 qualitative factor, suppose 2^6 and 2^{6-1} are too many runs. There is no 2^{6-2}_V design. Nonetheless, if we can live with a modelwith linear x and z terms plus x-z interaction, we can fit a 2^{6-2} factorial with defining relations $x_1x_2x_3 = I$, $x_1x_3x_4x_5z = I$, and $x_2x_4x_5z = I$. The design is resolution III but allows estimation of x_iz.

What about second order models? A safe, if costly, approach is to use a standard second order design for all combinations of the qualitative variables. An example of a more economical design is the following. Suppose we have two quantitative variables and one qualitative (two-level) variable. We can fit a model that is second order in the quantitative variables, and allows two-way interaction with the qualitative variable, with the following design

x_1	x_2	z
-1	-1	-1
1	-1	-1
$-\alpha$	0	-1
1	1	-1
0	α	-1
0	$-\alpha$	1
0	0	1
α	0	1
-1	1	1

This works because at each level of the qualitative variable we have a resolution III design, while taken across levels of the qualitative variable we can fit a full second-order model. Draper and John (1988) *Technometrics* have an extensive discussion of this sort of thing.

For more complicated designs, using computer generated designs base on optimality considerations is probably the best bet. As a grid of candidate point one can use the basic central composite design in the quantitative variables crossed with a complete factorial in the qualitative variables.

4. Incomplete Block Designs

As the name implies, incomplete block designs are arranged in blocks that are smaller than a complete replication. As one might imagine, many configurations of these designs can be developed. To introduce the basic concepts we will start with

5. Balanced Incomplete Block Designs

For example, suppose that we wish to conduct an experiment to compare 7 cooking oils with respect to taste. Because panelists cannot taste all oils, we decide that there is enough time to run 3 oils by each panelist. This then becomes an incomplete block design. Now what design do we use?

One way to set up this design is to take all possible combinations, i.e., $\binom{7}{2} =$ 35 panelists. Each pair of oils will occur with 5 panelists (given 2 oils, there are 5 other oils the pair can occur with).

An alternative to this design is one in which every pair of oils appears together with only one panelist. For example

Panelist	Oils
1	A,B,C
2	B,C,E
3	C,D,F
4	D,E,G
5	A,E,F
6	B,F,G
7	A,C,G

Both designs are balanced incoplete block designs because the number of times each pair of treatments occur together is the same. Balanced incomplete block (BIB) designs were first developed by Frank Yates in 1936.

They are designs for t treatments in b blocks. Each block contains k experimental units, or plots, with $k < t$. Furthermore, each treatment is replicated r times, i.e., appears in r replicates. We will assume that no treatment appears more than once in any block. Each pair of treatments occur in blocks λ times.

$b =$ number of blocks

$t =$ number of treatments

$r =$ number of reps of treatments

$k =$ number of plots per block

$\lambda =$ number of times each pair of treatments occurs

Each of these is an integer but of course they are not independent.

If n is the number of observations in the design then

$$n = rt = bk,$$

$$\lambda = \frac{r(k-1)}{t-1},$$

as the r blocks in which any particular treatment occurs contain $r(k-1)$ other plots, and these must be divided equally among the remaining $(t-1)$ treatments. Finally

$$b \geq t.$$

In the earlier example,

$$b = 7$$

$$t = 7$$

$$r = 3$$

$$k = 3$$

$$\lambda = \frac{r(k-1)}{t-1} = \frac{3(2)}{7-1} = 1$$

To illustrate some of the issues in the analysis of data collected in balanced incomplete block designs, consider the following data set.

Panelist	Oil and Response (7 pt scale)			Totals
1	A(7)	B(5)	D(5)	17
2	B(6)	C(4)	E(3)	13
3	C(5)	D(5)	F(2)	12
4	D(4)	E(4)	G(1)	9
5	A(6)	E(3)	F(2)	11
6	B(7)	F(4)	G(3)	14
7	A(6)	C(5)	G(4)	15
				91

$$\sum y^2 = 447$$
$$\sum y = 91$$
$$C.F. = (\sum y)^2/n = 394.33$$
$$SS_{TOT} = 52.67.$$

Oil Type	Treatment Totals (T)	B_t	Q	W
A	19	43	14	0
B	18	44	10	-10
C	14	40	2	-2
D	14	38	4	10
E	10	33	-3	24
F	8	37	-13	-8
G	8	38	-14	-14
	91	273	0	0

T = Treatment Totals

B_t = Total of all blocks in which the treatment occurs

$Q = kT - B_t$

$W = (t-k)T - (t-1)B_t + (k-1)n\bar{y}$

ANOVA

Source	df	SS*MS	F	
Total	20	52.67		
Treatment(adj)	6	32.86	5.48	7.50
Block(unadj)	6	14.00	2.33	
Error	8	5.81	0.73	

$$SS_{Tot} = \sum y^2 - C.F. = 52.67$$
$$C.F. = (\sum y)^2/n = 394.33$$
$$SS_{\text{Treatment(adj)}} = \frac{(t-1)\sum Q^2}{ktr(k-1)} = 32.86$$
$$SS_{\text{Blk(unadj)}} = \frac{17^2 + \cdots + 15^2}{3} - C.F. = 14.00$$
$$SS_{\text{Blk(adj)}} = SS_{\text{Blk(unadj)}} + SS_{\text{Treatment(adj)}} - SS_{\text{treatment(unadj)}} = 14.00 + 32.86 - 40.67 = 6.19$$

ANOVA

Source	df	SS	ms
Total	20	52.67	
Treatment(unadj)	6	40.67	6.78
Blk(adj)	6	6.19	$1.03 = E_b$
Error	8	5.81	$0.73 = E_e$

The adjusted block mean square is called the "inter block error," while the "error" is called the "intra block error."

A second type of incomplete block design is a

6. Partially Balanced Incomplete Block (PBIB)

The setup is the same as that which we discussed earlier for the balanced incomplete block design.

The major difference between the two designs is that the BIB has one associate class, while the PBIB has two associate classes. An associate class is a group of treatments such that they occur together within the same block the same number of times. As we saw earlier, a property of the BIB was that every pair of treatments occurred λ times, which means it has one associate class.

A PBIB has two associate classes. Consider the following example

Block	Treatments
1	1,2
2	1,3
3	2,3
4	3,4
5	3,5
6	4,5

1st associate ($\lambda_i = 1$)	2nd associate ($\lambda_2 = 0$)
1,2	1,4
1,3	1,5
2,3	2,4
3,4	2,5
3,5	
4,5	

This condition, however, is not sufficient for the design to be a PBIB.

A second condition is that every block contains k units. Every treatment occurs r times and no treatment appears more than once in a block.

Notice that in the design above the second condition is violated because treatment 3 occurs 4 times while all others appear twice. Therefore this is not a partially balanced incomplete block design.

Consider a second example

Blocks	Treatment
1	1,2
2	1,3
3	2,3
4	4,5
5	4,6
6	5,6

1st associate	2nd associate
1,2	1,4
1,3	1,5
2,3	1,6
4,5	2,4
4,6	2,5
5,6	2,6
	3,4
	3,5
	3,6

A third criterion for the design to be a partially balanced incomplete block design is examined with the following table.

		Treatment 1	
		1st assoc	2nd assoc
Treatment 2	1st assoc	3	
	2nd assoc		4,5,6

This pattern must be reproduced for any other pair of 1st associates.

Similarly for pairs of second associates

		Treatment 1	
		1st assoc	2nd assoc
Treatment 4	1st assoc		5,6
	2nd assoc	2,3	

This pattern must be reproduced for any other pair of 2nd associates.

Specifically, the 3rd criterion for a PBIB is "given any two treatments that are i-th associates, the number of treatments that j-th associates of the first treatment and k-th associates of the second treatment is the same no matter which pair of i-th associates we start with."

A more interesting example of a PBIB is

Block	Treatment
1	1,2,4,5
2	2,3,5,6
3	1,3,4,6
4	1,2,4,5
5	2,3,5,6
6	1,3,4,6

1st associate	2nd associate
1,4	1,2
2,5	1,3
3,6	1,5
	1,6
	2,3
6	2,4
	2,6
	3,5
	4,5
	4,6

		Treatment 1	
		1st assoc	2nd assoc
Treatment 4	1st assoc		
	2nd assoc		2,3,5,6

same for 2,5 and 3,6

		Treatment 1	
		1st assoc	2nd assoc
Treatment 2	1st assoc		5
	2nd assoc	4	3,6

same for all other pairs of 2nd associates