Stat 566
4-3-24

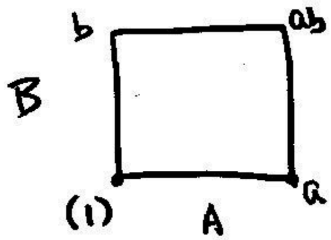
2^k factorial designs

1

k factors, each with 2 levels

The number 2^k tells how many runs are required for 1 replication of the experiment.

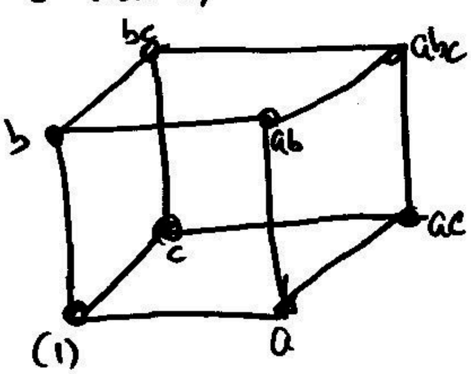
2^2 2 factors, each with 2 levels



2

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

2^3 3 factors, 2 levels each



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

Consider the 2^2 design, but with n replicates

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk} \qquad \begin{aligned} i &= 1,2 \\ j &= 1,2 \\ k &= 1, \dots, n \end{aligned}$$

Know: $\hat{\mu} = \bar{y}_{\cdots}$

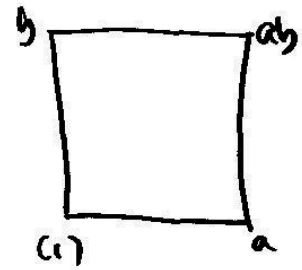
$\hat{\tau}_i = \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}$

Estimate $\tau_2 - \tau_1$

$$\hat{\tau}_2 - \hat{\tau}_1 = (\bar{y}_{2\cdot\cdot} - \bar{y}_{\cdots}) - (\bar{y}_{1\cdot\cdot} - \bar{y}_{\cdots})$$

5

$$= \bar{y}_{2\cdot\cdot} - \bar{y}_{1\cdot\cdot}$$



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

$$= \frac{1}{2n} \underbrace{\left[-(1) + a - b + ab\right]}_{C}$$

$$= \frac{C}{2n}$$

Note: C is a contrast

(6)

For a contrast of sums, $SS_C = \dfrac{C^2}{n \sum c_i^2}$

In our example, $SS_C = \dfrac{C^2}{n \cdot 4}$

In general, for a 2^k design with n replicates, the estimate of an effect due to one of the factors is $\dfrac{C}{2^{k-1} \cdot n}$

and $SS_C = \dfrac{C^2}{2^k \cdot n}$

(7)

df for 2^k with n reps.

$df_{TOT} = 2^k \cdot n - 1$

Each main effect has 1 df

Each interaction has 1 df

$$df_E = 2^k n - 1 - \underbrace{\left[\binom{k}{1} + \binom{k}{2} + \dots + \binom{k}{k} \right]}_{2^k - 1}$$

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

$$2^n = \sum_{k=0}^{n} \binom{n}{k}$$

8

$$df_E = 2^k n - 1 - (2^k - 1)$$
$$= 2^k n - 2^k = 2^k(n-1)$$

Note: If $n=1$, $df_E = 0$ & no F tests are possible.

What if no replication is possible?

1. Treat the highest-order interaction as negligible & remove it from the model.
   "Sparsity of effects" principle
   Gain: 1 df for error

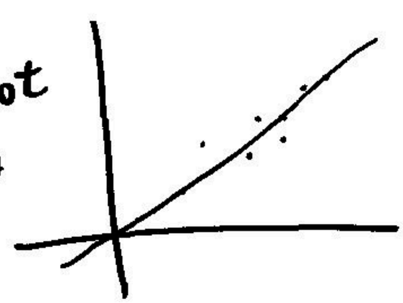
9

2. Rethink the entire design & pick a factor that you can live without.
   Delete it and all of its interactions.

   You now have a 2^{k-1} design with $n=2$

   Gain: 2^{k-1} df for error

3. Estimate all of the effects
   Plot on a normal probability plot



The effect estimates that are closest to the line are the ones least likely to be significant.

(10)

The "hierarchical principle" says that the model should not contain interactions unless all terms involved in the interaction are present in the model.

Gain: however many terms were removed

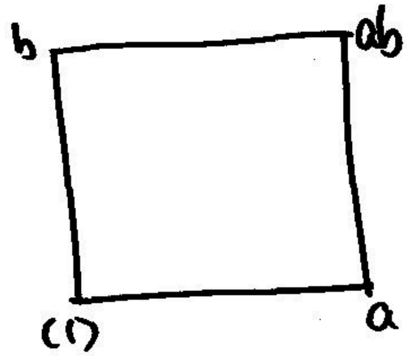
(4) Add several observations (n_E) at the center of the design (each factor is held at a level halfway between its high & low levels)

Gain: $n_E - 1$ df for error

(11)

2^K designs plus blocking

2^2 design



	I	A	B	AB
(1)	+	−	−	+
a	+	+	−	−
b	+	−	+	−
ab	+	+	+	+

Block 1	Block 2
(1)	a
ab	b

Note that the AB interaction is now confounded with blocks

2^3 design

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

Run in 2 blocks, with ABC confounded with blocks

(1)	a
ab	b
ac	c
bc	abc

↑ principal block

2^3 design in $\overset{2^{2}}{4}$ blocks

Pick 2 effects to be confounded with blocks

Select ABC & BC

+ −	+ +	− +	− −
(1)	a	b	ab
bc	abc	c	ac

Note: A is now confounded with blocks

Because $ABC \cdot BC = A$

Smarter selection: AB & BC $\Rightarrow$ AC

4.40. An engineer is studying the mileage performance characteristics of five types of gasoline additives. In the road test he wishes to use cars as blocks; however, because of a time constraint, he must use an incomplete block design. He runs the balanced design with the five blocks that follow. Analyze the data from this experiment (use $\alpha = 0.05$) and draw conclusions.

	Car				
Additive	**1**	**2**	**3**	**4**	**5**
1		17	14	13	12
2	14	14		13	10
3	12		13	12	9
4	13	11	11	12	
5	11	12	10		8

4.42. Seven different hardwood concentrations are being studied to determine their effect on the strength of the paper produced. However, the pilot plant can only produce three runs each day. As days may differ, the analyst uses the balanced incomplete block design that follows. Analyze the data from this experiment (use $\alpha = 0.05$) and draw conclusions.

Hardwood Concentration (%)	Days			
	1	**2**	**3**	**4**
2	114			
4	126	120		
6		137	117	
8	141		129	149
10		145		150
12			120	
14				136

Hardwood Concentration (%)	Days		
	5	**6**	**7**
2	120		117
4		119	
6			134
8			
10	143		
12	118	123	
14		130	127