Displays for Statistics 5303

Lecture 37

December 4, 2002

Christopher Bingham, Instructor

612-625-7023 (St. Paul)
612-625-1024 (Minneapolis)

Class Web Page
http://www.stat.umn.edu/~kb/classes/5303

More on confounding

For 2^3 factorial design, each effect is associated with a contrast:

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

You can use any contrast to define an incomplete block design with two blocks of size 4.

For example, the ABC contrast: Put all the -1 treatments in block I and all the +1 treatments in block II.
-1: (1), ab, ac, bc
+1: a, b, c, abc

This leads to the design

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

This is a 2^{3-1} design, a particular case of a 2^{k-p} design for g = 2^k treatments in b = 2^p blocks of size 2^{k-p}.

Blocks I has the treatments for which the ABC contrast coefficients are -1 and block II has the treatments for which the ABC contrast coefficients are +1.

ABC is the *defining contrast* and block I containing (1) is the *principal block*.

This just says which factor combinations are in which block. In using the design, you would randomly choose an actual block to be block I and randomly position the treatments in both blocks.

All you really need to get the whole design is the principal block I.

You can get the treatment combinations in block II by "multiplying" block I by a, dropping any squares

$a \circ (1) = \mathbf{a}$
$a \circ ab = a^2b = \mathbf{b}$,
$a \circ ac = a^2c = \mathbf{c}$
$a \circ bc = \mathbf{abc}$

In fact you get the same treatments when you multiply by b or c or abc, the other treatment combinations that are not in block I. For example,

$abc \circ (1) = \mathbf{abc}$
$abc \circ ab = a^2b^2c = \mathbf{c}$,
$abc \circ ac = a^2bc^2 = \mathbf{b}$
$abc \circ bc = ab^2c^2 = \mathbf{a}$

The ABC effect is *confounded* with blocks

The 3 factor block model is

$$y_{ijk\ell} = \mu + B_\ell + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} + \varepsilon_{ijk\ell}$$

where B_ℓ is a random or fixed block effect.

Then estimate of $\alpha\beta\gamma_{222}$ is

$$\widehat{\alpha\beta\gamma}_{222} = (-y_{111}+y_{211}+y_{121}-y_{221}+y_{112}-y_{212}-y_{122}+y_{222})/8$$
$$= \alpha\beta\gamma_{222} + (B_1 - B_2)/2 + \sum c_{ijk}^{ABC}\varepsilon_{ijk}$$

where $c_{ijk}^{ABC} = \pm 1/8$

$(B_1 - B_2)/2$ "contaminates" the estimate. When blocks are random, it greatly inflates the standard error. When they are fixed, it biases $\widehat{\alpha\beta\gamma}_{222}$

All the other contrasts associated with factorial effects are unconfounded with blocks because the contrasts are orthogonal to the ABC contrast.

$$\hat{\alpha}_2 = \alpha_2 + \sum c_{ijk}^{A}\varepsilon_{ijk}$$
$$\hat{\beta}_2 = \beta_2 + \sum c_{ijk}^{B}\varepsilon_{ijk}$$
. . . .
$$\widehat{\beta\gamma}_{jk} = \beta\gamma_{jk} + \sum c_{ijk}^{BC}\varepsilon_{ijk}, \text{ etc.}$$

No B_j's appear so these estimates are not confounded with block effects.

So for this design, all main effects and two-way interactions are unconfounded. Only the three-way interaction is confounded.

You can use any other factorial effect contrast to define blocks. Here is another 2^{3-1} design with defining contrast AB = (1, -1, -1, 1, 1 -1, -1, 1).

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

Now try to estimate $\alpha\beta_{22}$ using the AB contrast. You can check it has the form

$$\widehat{\alpha\beta}_{22} = (y_{111}-y_{211}-y_{121}+y_{221}+y_{112}-y_{212}-y_{122}+y_{222})/8$$
$$= \alpha\beta_{22} + (\beta_1 - \beta_2)/2 + \sum c_{ijk}^{AB}\varepsilon_{ijk}$$

Now the AB interaction is confounded, but no other is, not even ABC. A design based on A as defining contrast confounds the A main effect.

Finding confounded designs in MacAnova

For a 2^{k-p} design, macro **confound2()** finds the factor combinations in each block when you supply p defining contrasts.

2^{3-1} with defining contrast ABC

```
Cmd> confound2(vector(1,1,1)')
WARNING: searching for unrecognized macro confound2 near
confound2(
component: block1
(1) "(1)"
(2) "ab"
(3) "ac"
(4) "bc"
component: block2
(1) "a"
(2) "b"
(3) "c"
(4) "abc"
```

This confounds ABC = $A^1B^1C^1$. Note the "prime" (') in the argument. That makes it is a row vector (matrix with one row)

```
Cmd> confound2(vector(1,1,0)')  # defining contrast is AB
component: block1
(1) "(1)"
(2) "ab"
(3) "c"
(4) "abc"
component: block2
(1) "a"
(2) "b"
(3) "ac"
(4) "bc"
```

counfound3() works with 3^{k-p} designs.

Use **confound2()** to find the design for a 2^{4-1} design with defining contrast ABCD. There are $2^1 = 2$ blocks of size $2^3 = 8$. Only effect ABCD is confounded.

```
Cmd> confound2(vector(1,1,1,1)') # 2^(4-1)
component: block1
(1) "(1)"
(2) "ab"
(3) "ac"
(4) "bc"
(5) "ad"
(6) "bd"
(7) "cd"
(8) "abcd"
component: block2
(1) "a"
(2) "b"
(3) "c"
(4) "abc"
(5) "d"
(6) "abd"
(7) "acd"
(8) "bcd"
```

To confound ABCE, say, in a 2^5 design, use

```
confound2(vector(1,1,1,0,1)')
                 A B C D E
```

For a 2^{k-1} design, the argument is a 1 by k matrix or row vector. For a 2^{k-p} design, the argument must be a p by k matrix, with each row specifying a defining contrast

Statistics 5303 Lecture 37 December 4, 2002

What if block sizes of 2^{k-1} are too large.

For example, you are doing a 2^3 and want to use blocks of size 2^1, that is a 2^{3-2} design.

You need to use two defining contrasts. Use one to split in two groups of 4 and the second to split each group of 4 into two groups of 2

Let's try ABC and AB:

	(1)	a	b	ab	c	ac	bc	abc
ABC:	-1,	1,	1,	-1,	1	-1,	-1,	1
AB:	1,	-1,	-1,	1,	1	-1,	-1,	1

I	I
(1)	a
ab	b
ac	c
bc	abc

Blocks subdivided by AB

I_a	I_b	II_a	II_b
(1)	ac	c	a
ab	bc	abc	b

I_a and II_a contain treatments from I and II with AB = +1. I_b and II_b contain treatments from I and II with AB = -1

Statistics 5303 Lecture 37 December 4, 2002

What happens when you try to estimate C using contrast -1, -1, -1, -1, 1, 1, 1, 1?

$$\hat{\gamma}_2 = (-y_{111}-y_{211}-y_{121}-y_{221}+y_{112}+y_{212}+y_{122}+y_{222})/8$$

$$= \gamma_2 + (-B_1-B_3-B_3-B_1+B_4+B_2+B_2+B_4)/8 + \sum c_{ijk}{}^C\varepsilon_{ijk}$$

$$= \gamma_2 + (-B_1-B_3 + B_2+B_4)/4 + \sum c_{ijk}{}^C\varepsilon_{ijk}$$

So $\hat{\gamma}_2$ is contaminated with block effects so C is confounded with blocks.

Note: $AB \circ ABC = A^2B^2C = C$

This is an example of the important **general rule**:

When you confound any two contrasts, their generalized product is also confounded.

Statistics 5303 Lecture 37 December 4, 2002

Here **confound2()** finds the treatment combinations for each block.

```
Cmd> confound2(matrix(vector(1,1,1, 1,1,0),3)') # ABC and AB
component: block1
(1) "(1)"
(2) "ab"
component: block2
(1) "c"
(2) "abc"
component: block3
(1) "ac"
(2) "bc"
component: block4
(1) "a"
(2) "b"
```

Thus, with a confounded main effect, the choice of AB and ABC as defining contrasts is not a good one. You can use **choosedef2()** to find defining contrasts that are best in a certain sense.

```
Cmd> choosedef2(3,2,all:T) # For a 2^(3-2) design
component: generators
(1) "BC"
(2) "AB"
component: aberration
(1)          0          3          0
component: basis
(1,1)          0          1          1
(2,1)          1          1          0
               A          B          C
```

This gives defining contrasts AB and BC. This is better since $AB \circ BC = AB^2C = AC$. Main effects and ABC are not confounded. All 2-way interactions are confounded.

Component **generators** of **choosedef2()** output contains the defining contrasts as letters.

Component **basis** has the same information as a p by 2 matrix with columns corresponding to factors. 1 in a column means that factor is in the confounded effect. You can use **basis** as an argument to **choosedef2()**.

```
Cmd> basis <- choosedef2(3,2,all:T)$basis

Cmd> confound2(basis)
component: block1
(1) "(1)"
(2) "abc"
component: block2
(1) "ab"
(2) "c"
component: block3
(1) "a"
(2) "bc"
component: block4
(1) "b"
(2) "ac"
```

(1)	ab	a	b
abc	c	bc	ac

Component **aberration** tells you how many effects of different orders are confounded. You can tell 0 main effects 3 two-way interactions and 0 three-way interactions are confounded.

Let's try with 2^4 in blocks of 4: 2^{4-2}

```
Cmd> stuff <- choosedef2(4,2,all:T)

Cmd> stuff
component: generators
(1) "ABC"
(2) "ABD"
component: aberration
(1)          0          1          2          0
component: basis
(1,1)          1          1          1          0
(2,1)          1          1          0          1
```

This also confounds

$$ABC \circ ABD = A^2B^2CD = CD.$$

This is the one confounded two-way interaction.

A, AB, AC, AD, BC, BD, ACD, BCD and ABCD are unconfounded.

Here are the treatments:

```
Cmd> confound2(stuff$basis)
component: block1
(1) "(1)"
(2) "ab"
(3) "acd"
(4) "bcd"
component: block2
(1) "c"
(2) "abc"
(3) "ad"
(4) "bd"
component: block3
(1) "ac"
(2) "bc"
(3) "d"
(4) "abd"
component: block4
(1) "a"
(2) "b"
(3) "cd"
(4) "abcd"
```

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

Try a 2^{7-3}, timing how long it takes.

```
Cmd> timeit(stuff1 <- choosedef2(7,3,all:T);;)
Elapsed time is 10.8 seconds
```

This isn't very long but it's a noticeable wait.

```
Cmd> print(stuff1,format:"4.0f")
stuff1:
component: generators
(1) "ABDE"
(2) "ACDF"
(3) "BCDG"
component: aberration
(1)    0    0    0    7    0    0    0
component: basis
(1,1)    1    1    0    1    1    0    0
(2,1)    1    0    1    1    0    1    0
(3,1)    0    1    1    1    0    0    1
```

Q. What other effects are confounded in the 2^{7-3} design besides ACDE, BCDF, and ABCG spit out by **choosedef2()**?

A. All generalized products of these three:

$ACDE \circ BCDF = ABC^2D^2EF = ABEF$

$ACDE \circ ABCG = A^2BC^2DEG = BDEG$

$ACDE \circ BCDF = ABC^2D^2EF = ABEF$

and

$ACDE \circ BCDF \circ ABCG = A^2B^2C^3D^2EF = CGEF.$

This confirms **aberration** which says 7 four-way interactions are confounded, but no main effects, two-way, three-way, five-way, six-way, or seven-way interactions, and there are lots of unconfounded 4 way interactions.

Let's stretch it to 2^{8-3}:

```
Cmd> timeit(stuff2 <- choosedef2(8,3,all:T);;)
Elapsed time is 87.8 seconds

Cmd> print(stuff2,format:"3.0f")
stuff2:
component: generators
(1) "ABDEF"
(2) "ACDEG"
(3) "BCDEH"
component: aberration
(1)   0   0   0   3   4   0   0   0
component: basis
(1,1)   1   1   0   1   1   1   0   0
(2,1)   1   0   1   1   1   0   1   0
(3,1)   0   1   1   1   1   0   0   1
```

This took almost 90 seconds on a Macintosh G3 180. Much larger gets longer than one wants to wait.

You can save time by directing that **choosedef2()** not make an exhaustive search.

With **tries:m** instead f **all:T**, **choosedef2()** finds the best of **m** randomly selected sets of defining contrasts. It's can be much quicker, and although it isn't guaranteed to find the best it often does.

```
Cmd> timeit(stuff3 <- choosedef2(7,3,tries:1000);;)
Elapsed time is 2.5333 seconds

Cmd> print(abberation:stuff3$aberration,format:"4.0f")
aberration
(1)    0    0    0    7    0    0    0

Cmd> timeit(stuff4 <- choosedef2(8,3,tries:1000);;)
Elapsed time is 2.4 seconds

Cmd> print(abberation:stuff4$aberration,format:"4.0f")
abberation:
(1)    0    0    0    3    4    0    0    0
```

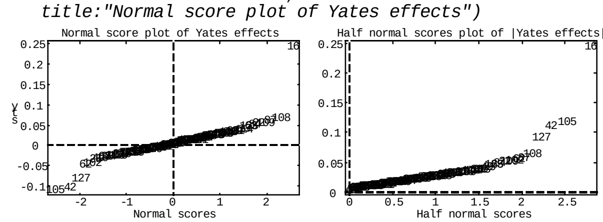
We got the same aberration patterns as with the exhaustive searches but both were much faster.

Here is a partial analysis of a 2^{7-4} design, 16 blocks of size 8.

```
Cmd> data <- read("","exmpl15.6")
exmpl15.6          128    9
) A data set from Oehlert (2000) \emph{A First Course in Design
) and Analysis of Experiments}, New York: W. H. Freeman.
)
) Table 15.6, p. 398
) Data for a 2^7 in standard order.  Factors are size of image,
) shape ofimage, color of image, orientation of image, duration
) of image vertical location of image, and horizontal location
) of image.
) ABCD, ACEG, BCE, BCFG, ACF, CDEF, ABG, BDEG, ADE, ADFG, BDF,
) EFG, CDG, ABEF, and ABCDEFG are confounded with blocks.
) Columns are block, A, B, C, D, E, F, G, and response
Read from file "TP1:Stat5303:Data:OeCh15.dat"

Cmd> makecols(data,block,a,b,c,d,e,f,g,y,factors:run(8))
Column 1 saved as factor block  with 16 levels
Column 2 saved as factor a  with 2 levels
Column 3 saved as factor b  with 2 levels
Column 4 saved as factor c  with 2 levels
Column 5 saved as factor d  with 2 levels
Column 6 saved as factor e  with 2 levels
Column 7 saved as factor f  with 2 levels
Column 8 saved as factor g  with 2 levels
Column 9 saved as vector y
```

Just as with non-confounded designs you can look at Yates effect's. Some, however are confounded and may be large because of large block differences.


```
Cmd> yts <- yates(y)

Cmd> chplot(rankits(yts),yts,xlab:"Normal scores",\
     title:"Normal score plot of Yates effects")

Cmd> chplot(halfnorm(abs(yts)),abs(yts),\
     xlab:"Half ormal scores",\
     title:"Normal score plot of Yates effects")
```



Effects 16, 105, 42 and 127 appear to be outliers. In binary notation these are

	GFEDCBA	Effect
16	= 0010000b	E
105	= 1101001b	ADFG
42	= 0101010b	BDF
127	= 1111111b	ABCDEFG

Note that the last three are all among the defining contrasts and represent differences between blocks, not treatment effects.

Here is an ANOVA including main effects plus 2-way, 3-way and 4-way interactions.

```
Cmd> anova("y=block + (a+b+c+d+e+f+g)^4")
Model used is y=block + (a+b+c+d+e+f+g)^4
WARNING: summaries are sequential
```

	DF	SS	MS
CONSTANT	1	30.44	30.44
block	15	1.3384	0.089225
a	1	0.02645	0.02645
b	1	0.014028	0.014028
c	1	0.0091125	0.0091125
d	1	0.0162	0.0162
e	1	1.9602	1.9602
f	1	0.0038281	0.0038281
g	1	0.0011281	0.0011281
a.b	1	0.005	0.005
a.c	1	0.0022781	0.0022781
a.d	1	0.029403	0.029403
a.e	1	0.0034031	0.0034031
a.f	1	0.017112	0.017112
a.g	1	0.01125	0.01125
b.c	1	0.021012	0.021012
b.d	1	0.0392	0.0392
b.e	1	5e-05	5e-05
b.f	1	0.0057781	0.0057781
b.g	1	0.0038281	0.0038281
c.d	1	0.0063281	0.0063281
c.e	1	0.00015312	0.00015312
c.f	1	0.0032	0.0032
c.g	1	0.0072	0.0072
d.e	1	0.0022781	0.0022781
d.f	1	0.00045	0.00045
d.g	1	0.02	0.02
e.f	1	0	0
e.g	1	0.00845	0.00845
f.g	1	0.00070312	0.00070312
a.b.c	1	0.0052531	0.0052531
a.b.d	1	0.0034031	0.0034031
a.b.e	1	0.0011281	0.0011281
a.b.f	1	0.00125	0.00125
a.b.g	0	0	undefined
a.c.d	1	0.0392	0.0392
a.c.e	1	0.0242	0.0242

a.c.f	0	0	undefined
a.c.g	1	0.0011281	0.0011281
a.d.e	0	0	undefined
a.d.f	1	3.125e-06	3.125e-06
a.d.g	1	0.0087781	0.0087781
a.e.f	1	0.045753	0.045753
a.e.g	1	0.0087781	0.0087781
a.f.g	1	0.0006125	0.0006125
b.c.d	1	0.024753	0.024753
b.c.e	0	0	undefined
b.c.f	1	0.037812	0.037812
b.c.g	1	0.0098	0.0098
b.d.e	1	0.0019531	0.0019531
b.d.f	0	0	undefined
b.d.g	1	0.0338	0.0338
b.e.f	1	0.0021125	0.0021125
b.e.g	1	0.00045	0.00045
b.f.g	1	0.025878	0.025878
c.d.e	1	0.01445	0.01445
c.d.f	1	0.00070312	0.00070312
c.d.g	0	0	undefined
c.e.f	1	0.00037812	0.00037812
c.e.g	1	0.067528	0.067528
c.f.g	1	0.027612	0.027612
d.e.f	1	0.015753	0.015753
d.e.g	1	0.0034031	0.0034031
d.f.g	1	0.0072	0.0072
e.f.g	0	0	undefined
a.b.c.d	0	0	undefined
a.b.c.e	1	0.0162	0.0162
a.b.c.f	1	0.015753	0.015753
a.b.c.g	1	0.00090313	0.00090313
a.b.d.e	1	0.0091125	0.0091125
a.b.d.f	1	0.0087781	0.0087781
a.b.d.g	1	0.0047531	0.0047531
a.b.e.f	0	0	undefined
a.b.e.g	1	0.038503	0.038503
a.b.f.g	1	0.00845	0.00845
a.c.d.e	1	0.0047531	0.0047531
a.c.d.f	1	0.00045	0.00045
a.c.d.g	1	0.0006125	0.0006125
a.c.e.f	1	0.0128	0.0128
a.c.e.g	0	0	undefined
a.c.f.g	1	0.0022781	0.0022781
a.d.e.f	1	0.00605	0.00605
a.d.e.g	1	0.012012	0.012012
a.d.f.g	0	0	undefined

a.e.f.g	1	0.011628	0.011628
b.c.d.e	1	0.02	0.02
b.c.d.f	1	0.00052812	0.00052812
b.c.d.g	1	0.014878	0.014878
b.c.e.f	1	0.0094531	0.0094531
b.c.e.g	1	0.0052531	0.0052531
b.c.f.g	0	0	undefined
b.d.e.f	1	0.00025313	0.00025313
b.d.e.g	0	0	undefined
b.d.f.g	1	0.0055125	0.0055125
b.e.f.g	1	0.032512	0.032512
c.d.e.f	0	0	undefined
c.d.e.g	1	0.0003125	0.0003125
c.d.f.g	1	0.1164	0.1164
c.e.f.g	1	0.023653	0.023653
d.e.f.g	1	0.0019531	0.0019531
ERROR1	28	0.59048	0.021088

The error term consists of all the pooled 5-, 6- and 7-way interaction SS.

Note the following:

- By far the largest effect mean square is for e.
- Terms a.b.g, a.c.f, a.d.e, b.c.e, b.c.e, b.d.f, c.d.g, e.f.g, a.b.c.d, a.b.e.f, a.c.e.g , a.d.f.g , b.c.f.g, b.d.e.g, c.d.e.f all have 0 degrees of freedom. These are the 15 confounded effects.