# ST 5203: Experimental Design

(Semester 1, AY 2017/2018)

**Text book:** *Experiments: Planning, Analysis, and Optimization (2nd. edition)*

by Jeff Wu and Mike Hamada

# Topic 5: Blocking and Confounding

- $2^k$ design in 2 blocks
- $2^3$ design in $2^2$ blocks
- $2^k$ design in $2^q$ blocks ($q < k$), determination of confounded effects
- Minimum aberration blocking scheme

# $2^2$ Design in 2 Blocks

- We start with the simplest situation: arrange a $2^2$ experiment in 2 blocks ($B1$,$B2$).
- Three possible blocking schemes are listed as follows

| Runs | $A$ | $B$ | $AB$ | Scheme 1 | Scheme 2 | Scheme 3 |
|---|---|---|---|---|---|---|
| (1) | $-$ | $-$ | $+$ | $B1$ | $B1$ | $B1$ |
| $a$ | $+$ | $-$ | $-$ | $B1$ | $B2$ | $B1$ |
| $b$ | $-$ | $+$ | $-$ | $B2$ | $B1$ | $B2$ |
| $ab$ | $+$ | $+$ | $+$ | $B2$ | $B2$ | $B1$ |

- Which scheme is better and why?

# Which Scheme is Better and Why

- Remember the estimates of effects:

$$\begin{aligned}
\mu &= \frac{(1)+a+b+ab}{4} \\
A &= \frac{ab+a-b-(1)}{2} \\
B &= \frac{ab+b-a-(1)}{2} \\
AB &= \frac{ab+(1)-a-b}{2}
\end{aligned}$$

- Scheme 1: $A$ and $AB$ are valid, but $B$ is confused (confounded) with the block.
- Scheme 2: $B$ and $AB$ are valid, but $A$ is confused (confounded) with the block.
- Scheme 3: $A$, $B$ and $AB$ are all inappropriately estimated (confounded). (Reason: check the orthogonality of two contrast vectors.)
- Is there any better way for blocking?

# Improved Blocking Scheme and Reasoning

- Consider scheme 4: $B2, B1, B1, B2$. Now, main effects $A$, $B$ are valid, however, $AB$ is confounded with the blocking.
- Scheme 2 is using the column of $A$ as the blocking scheme, "$-$" is used as Block 1 "$+$" as Block 2. Thus, $A$ is confounded with the blocking. Similarly for Scheme 1 (for $B$) and Scheme 4 (for $AB$).
- The conclusion above does not happen by chance. Usually, in a $2^k$ design with 2 blocks, if we choose one column of the model matrix as the blocking scheme, the corresponding effect is sacrificed (confounded with the blocking), but all the other effects are estimated appropriately.
- See the example of $2^3$ design in 2 blocks in the next slide.

# Schemes of $2^3$ Design in 2 Blocks

| Trt | $A$ | $B$ | $C$ | $AB$ | $AC$ | $BC$ | $ABC$ | Scheme 1 | Scheme 2 |
|---|---|---|---|---|---|---|---|---|---|
| (1) | $-$ | $-$ | $-$ | $+$ | $+$ | $+$ | $-$ | 1 | 1 |
| $a$ | $+$ | $-$ | $-$ | $-$ | $-$ | $+$ | $+$ | 2 | 2 |
| $b$ | $-$ | $+$ | $-$ | $-$ | $+$ | $-$ | $+$ | 1 | 2 |
| $ab$ | $+$ | $+$ | $-$ | $+$ | $-$ | $-$ | $-$ | 1 | 1 |
| $c$ | $-$ | $-$ | $+$ | $+$ | $-$ | $-$ | $+$ | 2 | 2 |
| $ac$ | $+$ | $-$ | $+$ | $-$ | $+$ | $-$ | $-$ | 2 | 1 |
| $bc$ | $-$ | $+$ | $+$ | $-$ | $-$ | $+$ | $-$ | 2 | 1 |
| $abc$ | $+$ | $+$ | $+$ | $+$ | $+$ | $+$ | $+$ | 1 | 2 |

## Schemes of $2^3$ Design in 2 Blocks (Cont.)

- Scheme 1 does not use any effect as the blocking reference. If we check results,
  - $A$, $BC$, $ABC$ are un-confounded.
  - $B$, $C$, $AB$, $AC$ are confounded.
- Clearly, scheme 2 is using $ABC$ column as the blocking reference. After checking the results, we found that only $ABC$ is confounded with the blocking factor. All the other effects are not confounded.
- Scheme 2 outperforms scheme 1.

# Schemes of $2^3$ Design in 2 Blocks (Cont.)

- Due to blocking, some effects suppose to be confounded with the blocking. A good scheme is to minimize the number of confounded effects.
- Besides, refer to the fundamental principles in factorial experiments, we should try to confound higher order effects instead of lower order ones, because it is preferable to sacrifice less important effects.
- Based on arguments above, for $2^3$ design, Scheme 2 is the best blocking scheme without further scientific knowledge on the different effects.
- What is the best blocking scheme for a general $2^k$ design in 2 blocks, if we don't have any (prior) knowledge about which effect is significant in real practice?

# Explanation on Blocking: From Mathematical Point of View

- Recall the regression equations for a $2^3$ design:

$$\begin{aligned} y &= \mu + \frac{A}{2}x_1 + \frac{B}{2}x_2 + \frac{C}{2}x_3 + \frac{AB}{2}x_1x_2 + \frac{AC}{2}x_1x_3 \\ &\quad + \frac{BC}{2}x_2x_3 + \frac{ABC}{2}x_1x_2x_3 \end{aligned}$$

  where $x_i = \pm 1$ for different runs. In total, there are 8 equations for the $2^3$ full factorial design.
  (Note that we don't consider the error terms here, since we don't have enough data.)

- Without blocking, all the 8 equations above are sharing the same $\mu$, i.e. all the runs have the same grand mean.

- When we divide the experiment into two blocks, then we will have two grand means: $\mu_1$ for block 1, $\mu_2$ for block 2.

# Regression Equations of $2^3$ Design with Blocking

$$
\begin{aligned}
(1) &= \mu_1 - \frac{A}{2} - \frac{B}{2} - \frac{C}{2} + \frac{AB}{2} + \frac{AC}{2} + \frac{BC}{2} - \frac{ABC}{2} && (1)\\
a &= \mu_2 + \frac{A}{2} - \frac{B}{2} - \frac{C}{2} - \frac{AB}{2} - \frac{AC}{2} + \frac{BC}{2} + \frac{ABC}{2} && (2)\\
b &= \mu_2 - \frac{A}{2} + \frac{B}{2} - \frac{C}{2} - \frac{AB}{2} + \frac{AC}{2} - \frac{BC}{2} + \frac{ABC}{2} && (3)\\
ab &= \mu_1 + \frac{A}{2} + \frac{B}{2} - \frac{C}{2} + \frac{AB}{2} - \frac{AC}{2} - \frac{BC}{2} - \frac{ABC}{2} && (4)\\
c &= \mu_2 - \frac{A}{2} - \frac{B}{2} + \frac{C}{2} + \frac{AB}{2} - \frac{AC}{2} - \frac{BC}{2} + \frac{ABC}{2} && (5)\\
ac &= \mu_1 + \frac{A}{2} - \frac{B}{2} + \frac{C}{2} - \frac{AB}{2} + \frac{AC}{2} - \frac{BC}{2} - \frac{ABC}{2} && (6)\\
bc &= \mu_1 - \frac{A}{2} + \frac{B}{2} + \frac{C}{2} - \frac{AB}{2} - \frac{AC}{2} + \frac{BC}{2} - \frac{ABC}{2} && (7)\\
abc &= \mu_2 + \frac{A}{2} + \frac{B}{2} + \frac{C}{2} + \frac{AB}{2} + \frac{AC}{2} + \frac{BC}{2} + \frac{ABC}{2} && (8)
\end{aligned}
$$

# Effects Solvable or Un-solvable

- In total, we have 8 equations, but with 9 parameters. Generally speaking, this system of equations is not solvable. Fortunately, if we carefully check the equations and parameters, some parameters are solvable.
- For example,
  - Add up equations (2), (4), (6) and (8): $a + ab + ac + abc = 2\mu_1 + 2\mu_2 + 2A$;
  - Add up equations (1), (3), (5) and (7): $(1) + b + c + bc = 2\mu_1 + 2\mu_2 - 2A$.

  Thus, $A$ can be solved by the subtraction of the two equations above. Similarly, all the other effects except $ABC$ can be solved.
- Check the effect $ABC$ in the equation. "$+ABC$" is purposely matched with $\mu_2$ and "$-ABC$" with $\mu_1$. This is why $ABC$ is unsolvable.

# Rules of Multiplication

- Recall that the uppercase letters (such as $A, B, AB$) have several meanings, one of which is the contrast vector.
- Here, the operation of multiplication between two or more uppercase letters means the element-wise product of their contrast vectors.
- For example, in the table $A = (-1, 1, -1, 1, -1, 1, -1, 1)^\top$, $B = (-1, -1, 1, 1, -1, -1, 1, 1)$. Therefore, $D2 = AB = (1, -1, -1, 1, 1, -1, -1, 1)^\top$.
- Define $\boldsymbol{I} = (1, 1, \ldots, 1)^\top$. $\boldsymbol{I}$ is the identity element. Under this definition of multiplication, the product of any effect with itself is always $I$. For example, $A^2 = A \times A = \boldsymbol{I}$, $B \times \boldsymbol{I} = B$.

# $2^3$ Design in 4 Blocks

- For 2 blocks, we need to use one effect column as blocking scheme: $-$, Block 1; $+$, Block 2.
- If we want to perform the experiment in 4 blocks, let us try to check two effect columns. Consider the following scheme:

| Effect 1 ($\boldsymbol{B}_1$) | Effect 2($\boldsymbol{B}_2$) | Blocks |
|---|---|---|
| $-$ | $-$ | $B1$ |
| $+$ | $-$ | $B2$ |
| $-$ | $+$ | $B3$ |
| $+$ | $+$ | $B4$ |

- Thus, two effect columns can totally determine a scheme with 4 blocks. There are 3 block effects: $\boldsymbol{B}_1$, $\boldsymbol{B}_2$, and $\boldsymbol{B}_1\boldsymbol{B}_2$ (explain). The interactions like $\boldsymbol{B}_1\boldsymbol{B}_2$ are called generalized interactions.

# $2^3$ Design in 4 Blocks (Cont.)

The following is an example of $2^3$ design in 4 blocks, with $ABC$ and $AB$ as the blocking factors.

| Trt | $A$ | $B$ | $C$ | $AC$ | $BC$ | $\boldsymbol{B}_1 = ABC$ | $\boldsymbol{B}_2 = AB$ | Scheme 1 |
|---|---|---|---|---|---|---|---|---|
| (1) | $-$ | $-$ | $-$ | $+$ | $+$ | $-$ | $+$ | $B3$ |
| $a$ | $+$ | $-$ | $-$ | $-$ | $+$ | $+$ | $-$ | $B2$ |
| $b$ | $-$ | $+$ | $-$ | $+$ | $-$ | $+$ | $-$ | $B2$ |
| $ab$ | $+$ | $+$ | $-$ | $-$ | $-$ | $-$ | $+$ | $B3$ |
| $c$ | $-$ | $-$ | $+$ | $-$ | $-$ | $+$ | $+$ | $B4$ |
| $ac$ | $+$ | $-$ | $+$ | $+$ | $-$ | $-$ | $-$ | $B1$ |
| $bc$ | $-$ | $+$ | $+$ | $-$ | $+$ | $-$ | $-$ | $B1$ |
| $abc$ | $+$ | $+$ | $+$ | $+$ | $+$ | $+$ | $+$ | $B4$ |

# Properties of Scheme 1 and Possible Improvement

- Clearly, since we are using $AB$ and $ABC$ as the blocking scheme, they are confounded with the blocking effects.
- Is there any other effect lost in this scheme? Unfortunately, yes. The main effect $C$ is also confounded, because $\boldsymbol{B}_1\boldsymbol{B}_2 = ABCAB = C$.
- Let us try a different Scheme 2, in which $\boldsymbol{B}_1 = AC$, $\boldsymbol{B}_2 = AB$. Now, both $AC$ and $AB$ are confounded with block effects. Besides, the effect $BC$ is also confounded, because $\boldsymbol{B}_1\boldsymbol{B}_2 = ACAB = BC$. All other effects (including all main effects and $ABC$) are not confounded.
- How do we identify all confounded effect(s) in general?

## $2^k$ Design in $2^q$ Blocks

- In general, consider a $2^k$ design in $2^q$ blocks ($q < k$).
- $q$ effects are required as the coding effects.
- The confounded effects are:

  $\binom{q}{1}$ coding effects,

  $\binom{q}{2}$ multiplications of any two coding effects,

  ......,

  $\binom{q}{q}$ multiplications of all $q$ coding effects,

  Beside, $\binom{q}{0} = 1$ grand mean is also confounded.

Here $\binom{n}{m} = \dfrac{n!}{m!(n-m)!}$ is the combinatorial number of "$n$ choose $m$" (also called the binomial coefficient).

## $2^k$ Design in $2^q$ Blocks (Cont.)

- First choose $q$ different effects $v_1, \ldots, v_q$.
- Define the confounding relations $\boldsymbol{B}_1 = v_1$, $\boldsymbol{B}_2 = v_2$, ..., $\boldsymbol{B}_q = v_q$.
- Take the 2-way, 3-way, ..., $q$-way products of them: $\boldsymbol{B}_1\boldsymbol{B}_2 = v_1v_2, \ \ldots \ , \boldsymbol{B}_{q-1}\boldsymbol{B}_q = v_{q-1}v_q, \ \boldsymbol{B}_1\boldsymbol{B}_2\boldsymbol{B}_3 = v_1v_2v_3, \ \ldots \ , \boldsymbol{B}_1 \cdots \boldsymbol{B}_q = v_1 \cdots v_q$.
- In total, there are $2^q - 1$ possible products of $\boldsymbol{B}$'s. They and the identity element $\boldsymbol{I}$ together form a group, called the block defining contrast subgroup.
- The $2^q - 1$ effects of $v_1, \ldots, v_q, v_1v_2, \ldots, v_1 \cdots v_q$ are all the effects that are confounded with block effects.

## How to Choose the Best Blocking Scheme?

- In the previous example of $2^3$ design with $2^1$ blocks, we prefer the scheme with $\boldsymbol{B} = ABC$ over the scheme with $\boldsymbol{B} = AB$.
- Based on the effect hierarchy principle, in the previous example of $2^3$ design with $2^2$ blocks, we prefer the scheme with $\boldsymbol{B}_1 = AB, \boldsymbol{B}_2 = AC, \boldsymbol{B}_1\boldsymbol{B}_2 = BC$ over the scheme with $\boldsymbol{B}_1 = ABC, \boldsymbol{B}_2 = AB, \boldsymbol{B}_1\boldsymbol{B}_2 = C$, because the latter scheme has a main effect $C$ confounded while the former one does not.
- In general, we need a systematic way to choose the best blocking scheme.

## Minimum Aberration Blocking Scheme

Consider the $2^k$ design with $2^q$ blocks.

- If $b$ represents a blocking scheme, define $g_i(b)$ to be the number of $i$-factor interactions that are confounded with block effects, under the scheme $b$.
- We check $g_i(b)$ from $i = 1, 2, ..., k$. $\sum_{i=1}^{k} g_i(b) = 2^q - 1$.
- For two blocking schemes $b_1$ and $b_2$, define $r$ to be the smallest integer such that $g_r(b_1) \neq g_r(b_2)$.
- The scheme $b_1$ is said to have less aberration than the scheme $b_2$, if $g_r(b_1) < g_r(b_2)$.
- A blocking scheme is said to have the minimum aberration, if there is no other block scheme with less aberration.

## Minimum Aberration Blocking Scheme (Cont.)

- In the previous example of $2^3$ design with $2^2$ blocks, the scheme $b_1$ with $\boldsymbol{B}_1 = AB, \boldsymbol{B}_2 = AC, \boldsymbol{B}_1\boldsymbol{B}_2 = BC$ has $g_1(b_1) = 0$, while the scheme $b_2$ with $\boldsymbol{B}_1 = ABC, \boldsymbol{B}_2 = AB, \boldsymbol{B}_1\boldsymbol{B}_2 = C$ has $g_1(b_2) = 1$. Therefore, $b_1$ has less aberration than $b_2$. In fact, $b_1$ is the minimum aberration blocking scheme for $k = 3, q = 2$.

- We can consider another example of $2^4$ design with $2^2$ blocks.
  - Scheme $b_1$: $\boldsymbol{B}_1 = ABC$, $\boldsymbol{B}_2 = ABCD$.
  - Scheme $b_2$: $\boldsymbol{B}_1 = AB$, $\boldsymbol{B}_2 = CD$.
  - Scheme $b_3$: $\boldsymbol{B}_1 = ABC$, $\boldsymbol{B}_2 = ABD$.

- Check that $g_1(b_1) = 1$, $g_1(b_2) = 0$, $g_1(b_3) = 0$, $g_2(b_2) = 2$, $g_2(b_3) = 1$. So Scheme $b_3$ has less aberration than Schemes $b_1$ and $b_2$. In fact, Scheme $b_3$ is the minimum aberration blocking scheme for $k = 4, q = 2$.