16.0 More on Fractional Factorial Designs

- Answer Questions; Review
- Fractional Factorials
- Example
- More Exotic Designs
16.1 Review of Factorial Designs

Today’s lecture extends our previous discussion of factorial designs. To ensure narrative continuity, we start with a quick review.

The following table is a full $2^3$ factorial design. The signs in each interaction column are found by multiplying the signs in corresponding main-effect columns.
<table>
<thead>
<tr>
<th>run</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>AB</th>
<th>AC</th>
<th>BC</th>
<th>ABC</th>
<th>obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>$Y_{111}$</td>
</tr>
<tr>
<td>2</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>$Y_{211}$</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>$Y_{121}$</td>
</tr>
<tr>
<td>4</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>$Y_{221}$</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>$Y_{112}$</td>
</tr>
<tr>
<td>6</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>$Y_{212}$</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>$Y_{122}$</td>
</tr>
<tr>
<td>8</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>$Y_{222}$</td>
</tr>
</tbody>
</table>
The main effect due to factor A is the average difference between the high and low levels of factor A, or:

\[ \alpha = \frac{1}{2} \left( \frac{Y_1 + Y_4 + Y_6 + Y_8}{4} - \frac{Y_1 + Y_3 + Y_5 + Y_7}{4} \right) \]

and similarly for the main effects of B and C. So, as per the signs in the table,

\[ \alpha = \frac{1}{8}(-Y_1 + Y_2 - Y_3 + Y_4 - Y_5 + Y_6 - Y_7 + Y_8). \]

The AB interaction is half the difference between the main effect of factor A at the high level of factor B and that at the low level of factor B, or:

\[ (\alpha\beta) = \frac{1}{2} \left[ \frac{1}{2} \left( \frac{Y_4 + Y_8}{2} - \frac{Y_3 + Y_7}{2} \right) - \frac{1}{2} \left( \frac{Y_2 + Y_6}{2} - \frac{Y_1 + Y_5}{2} \right) \right] \]

As per the signs in the table,

\[ (\alpha\beta) = \frac{1}{8}(Y_1 - Y_2 - Y_3 + Y_4 + Y_5 - Y_6 - Y_7 + Y_8). \]
The ABC interaction is half the difference between the two-factor interaction AB at the high and low levels of factor C, or

\[
(\alpha/\beta/\gamma) = \frac{1}{2} \left[ \frac{1}{2} \left( \frac{Y_8 - Y_7}{2} - \frac{Y_6 - Y_5}{2} \right) - \frac{1}{2} \left( \frac{Y_4 - Y_3}{2} - \frac{Y_2 - Y_1}{2} \right) \right]
\]

Note that

\[
(\alpha/\beta/\gamma) = \frac{1}{8} (-Y_1 + Y_2 + Y_3 - Y_4 + Y_5 - Y_6 - Y_7 + Y_8)
\]
as per the signs in the previous table.

These relationships give us an automatic way to calculate the main effects (and, implicitly, the sums of squares) for inference.
But $2^k$ observations gets expensive. Instead, one can carefully select half that number so as to still permit estimation of main effects and low-order interactions.
Consider the $2^{3-1}$ fractional factorial design (The previous figure gave two illustrations.)

<table>
<thead>
<tr>
<th>run</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>AB</th>
<th>AC</th>
<th>BC</th>
<th>ABC</th>
<th>obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>4</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Note that because we have taken only half of the 8 observations needed for a full factorial, some of the columns have identical entries.

Columns that have identical entries correspond to effects that are confounded or aliased.
In order to estimate the effects, note that:

\[
\frac{1}{4}(Y_1 + Y_2 + Y_3 + Y_4) = \mu + (\alpha\beta\gamma)
\]

\[
\frac{1}{4}(-Y_1 + Y_2 - Y_3 + Y_4) = \alpha + (\beta\gamma)
\]

\[
\frac{1}{4}(-Y_1 - Y_2 + Y_3 + Y_4) = \beta + (\alpha\gamma)
\]

\[
\frac{1}{4}(Y_1 - Y_2 - Y_3 + Y_4) = \gamma + (\alpha\beta)
\]

Thus the estimate of the mean is confounded with the three-way interaction, the estimate of the A effect is confounded with the BC interaction, the estimate of the B effect is confounded with the AC interaction, and the estimate of the C effect is confounded with the AB interaction.

If one assumes that there are no interactions, then one can make tests about the main effects, or use a normal probability plot.
Note that we write $2^{k-p}$ to denote a fractional factorial design in which each factor has 2 levels, there are $k$ factors, and we are taking a $1/2^p$ fraction of the number of possible factor level combinations.

In order to construct a fractional factorial that deliberately confounds pre-selected factors, one needs to use a **generator**.

The generator uses the fact that squaring the entries in any given column gives a column of ones, which can be thought of as an identity element $I$. If we want to confound the A effect with the BC interaction, then that is equivalent to declaring $A \ast BC = ABC = I$. It follows that $B = BI = B \ast ABC = AC$, so $B$ is confounded with the AC interaction. Similarly, $C$ is confounded with $AB$, and the overall mean ($I$) is confounded with $ABC$.

Here, the generator is the relationship $I=ABC$. 
Consider a $2^{4-1}$ fractional factorial design.

<table>
<thead>
<tr>
<th>run</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>AB</th>
<th>AC</th>
<th>BC</th>
<th>ABC</th>
<th>obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>4</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>6</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

$Y_{1111}$, $Y_{2112}$, $Y_{1212}$, $Y_{2211}$, $Y_{1122}$, $Y_{2121}$, $Y_{1221}$, $Y_{2222}$
What factors and interactions are confounded?

First, note that the ABC interaction has the same signs in its column as D. Thus D and ABC are confounded, or D=ABC. This implies that I=ABCD is the generator. (Why?)

> From this, the confounding pattern is as follows:

- A is confounded with BCD
- B is confounded with ACD
- C is confounded with ABD
- D is confounded with ABC
- AB is confounded with CD
- AC is confounded with BD
- AD is confounded with BC
- ABCD is confounded with $\mu$
One can go the other way, picking the generator and then deriving the confounding pattern (and thus the design). For example, suppose we had decided to confound D with AB.

In that case, the generator is \( I=ABD \). Thus:

- A is confounded with BD
- B is confounded with AD
- C is confounded with ABCD
- D is confounded with AB
- AC is confounded with BCD
- ACD is confounded with BC
- ABC is confounded with CD
- ABD is confounded with \( \mu \).
The first design has **Resolution IV**. This is because all the main effects (order I) are confounded with three-way interactions (order III) and the resolution is I+III = IV. Such a design is sometimes written as $2_{IV}^{4-1}$.

The second design has **Resolution III**. This is because one of the main effects (order I) is confounded with a two-way interaction (order II), giving resolution I+II = III. This design is written as $2_{III}^{4-1}$.

In general, we prefer designs that have higher resolution. This ensures that one can make relatively clean tests of main effects (and, for larger numbers of factors, the tests of two-way interactions).
Consider a $2^{4-1}$ fractional factorial design. The data are from a slump test of concrete; the four factors are water (A), mixing (B), gravel proportion (C), and cement proportion (D).

<table>
<thead>
<tr>
<th>run</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>AB</th>
<th>AC</th>
<th>BC</th>
<th>ABC</th>
<th>obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>23</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>27</td>
</tr>
<tr>
<td>4</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>24</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>23</td>
</tr>
<tr>
<td>6</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>33</td>
</tr>
<tr>
<td>8</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>37</td>
</tr>
</tbody>
</table>
If the null hypothesis of no effects is true, then the 3.5 and 3.0 look relatively large. These deserve more study.
16.2 Higher-Order Fractions

Consider an experiment designed to study seven factors in only eight runs. This means we need a $2^{7-4}$ fractional factorial design.

To do this, one needs more than one generator (in fact, one needs four generators, since each halves the number of observations). One strategy is to write out a full $2^3$ factorial design, and then associate (confound or alias) the interactions with each of the four additional factors.
This implies \( D=AB, \ E=AC, \ F=BC, \) and \( G=ABC, \) from which we get the generating relations \( I=ABD, \ I=ACE, \ I=BCF, \) and \( I=ABCG. \)
Since the product of I with itself is also I, then one can multiply any pair or triple or quadruple of the generating relations and still get I. Thus:

\[
\begin{align*}
I &= ABD = ACE = BCF = ABCG \\
   &= BCDE = ACDF = CDG = ABEF \\
   &= BEG = AFG = DEF = ADEG \\
   &= BDFG = CEFG = ABCDEFG
\end{align*}
\]

This is cumbersome and tedious, but it enables one to calculate all the aliased main effects and interactions.

This is called the **defining relation**.
If we multiply the defining relation by A, we find that

\[ A = BD = CE = ABCF = BCG \]
\[ = ABCDE = CDF = ACDG = BEF \]
\[ = ABEG = FG = ADEF = DEG \]
\[ = ABDFG = ACEFG = BCDEFG \]

Thus A is confounded with two-way, three-way, and higher interactions.

Similar calculations can be done for the other main effects. Each main effect is confounded with a two-way interaction (e.g., G is confounded with CD, BE, and AF. This means we have a \(2^{7-4}_{III}\) design.
Box and Hunter (Technometrics, 1961, 311-351) describe a $2^{7-4}_{III}$ design for a company that experienced difficulty (long delays) with a filtration system that was part of their manufacturing process. The factors were:

A: water supply (town, well)

B: raw material (two suppliers)

C: temperature at filtration (high and low)

D: recycling (included or omitted)

E: rate of addition of caustic soda (slow, fast)

F: type of filter cloth (old, new)

G: prior hold-up time (short, long)

The raw data from the experiment are shown in the following table.
If one estimates the seven main effects (each of which is aliased with interactions), one finds $A$ is -5.4, $B$ is -1.4, $C$ is -8.3, $D$ is 1.6, $E$ is -11.4, $F$ is -1.7, and $G$ is 0.3. It seems that $E$, $C$, and $A$ (or corresponding interactions) merit more thought.
These extreme fractions are difficult—one cannot get much information from, say, eight observations, but these designs cleverly ensure that one gets the most one can.

The designs are especially useful in stagewise experimentation. Often one starts with a large set of factors, and then develops a series of fractional factorial experiments that home in on the most important effects.

In the filtration experiment, the firm can now design follow-up experiments that focus on the key factors. Or they can try to run their system at the high levels of A, B, C, D, E, F, and G, which gave a filtration time of 38.7.
16.4 Exotic Designs

We have not talked about:

- Balanced Incomplete Block Designs
- Partially Balanced Incomplete Block Designs with k Associate Classes
- $p^{k-r}$ designs, or $p^{k-r}q^{k-s}$ designs, and other generalizations
- Plackett-Burman designs, and other methods for response surface analysis
- D-optimality, G-optimality, E-optimality, and so forth.
- Repeated measures designs (MANOVA)
- ANACOVA