Chapter 19 - Factorial Designs

Factorial design - all combinations of levels of (two) factors are observed.

Balanced design - all combinations have the same number of observations.

Why run a factorial design? Why not optimize the two factors, one at a time?

Model

\[ ab = (\# \text{ of combinations of levels}) \]
**Example:** Student achievement at a learning task

Minutes in Isolation

<table>
<thead>
<tr>
<th>Encouragement</th>
<th>None</th>
<th>M₁₁</th>
<th>M₁₂</th>
<th>M₁₃</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Verbal</td>
<td>M₂₁</td>
<td>M₂₂</td>
<td>M₂₃</td>
</tr>
</tbody>
</table>

**Example:** Orange tree trunk diameter changes

Calcium Level

<table>
<thead>
<tr>
<th>A Soil PH</th>
<th>100</th>
<th>200</th>
<th>300</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>M₁₁</td>
<td>M₁₂</td>
<td>M₁₃</td>
</tr>
<tr>
<td>5</td>
<td>M₂₁</td>
<td>M₂₂</td>
<td>M₂₃</td>
</tr>
<tr>
<td>6</td>
<td>M₃₁</td>
<td>M₃₂</td>
<td>M₃₃</td>
</tr>
</tbody>
</table>

Models for cell means:

Main Effects Model
With this model it is easy to interpret the result of changing the level of any given factor:

\[ \mu_{ij} - \mu_{i'j'} = (\mu_{..} + \alpha_i + \beta_j) - (\mu_{..} + \alpha_{i'} + \beta_{j'}) \]

\[ = \beta_j - \beta_{j'} \]

and

\[ \mu_{ij} - \mu_{i'j} = \alpha_i - \alpha_{i'} \]

Here

Interaction Model

\[ \mu_{..} = \sum_{i,j} \mu_{ij}/ab \]

\[ \mu_{ij} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} \]
Now the effect of changing a level of one factor depends on the level of the other.

Model for Data:

\[ Y_{ijk} = \mu_{ij} + E_{ijk} \]

The \( E_{ijk} \)'s are all iid \( N(0, \sigma^2) \).

Notation:

\[ \bar{Y}_{ij} = \frac{\sum_{k=1}^{n} Y_{ijk}}{n} = \text{Ave. of } i,j^{th} \text{ cell obs} \]
\[ \hat{\mu}_{..} = \bar{Y}_{..} \quad \text{estimates } \mu_{..} \]

\[ \hat{\alpha}_{i} = \bar{Y}_{i..} - \bar{Y}_{..} \quad \text{estimates main effect of } i^{th} \text{ level of factor } A \]

---

**Exercise:** Show that

\[ \sum_{i=1}^{a} \hat{\alpha}_{i} = 0 = \sum_{j=1}^{b} \hat{\beta}_{j} \]

\[ \sum_{i=1}^{a} (\times \beta)_{ij} = 0 = \sum_{j=1}^{b} (\times \beta)_{ij} \]

---

For this model, we want to determine whether the interactions or the main effects are significant.

**Important Point:** If interactions are present, then main effects are no longer interpretable.
\[ \text{SSTO} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (Y_{ijk} - \bar{Y}_{..})^2 \]

\[ = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \left[ (Y_{ijk} - \bar{Y}_{ij}) + (\bar{Y}_{ij} - \bar{Y}_{..} - \bar{Y}_{ij} + \bar{Y}_{...}) \right. \]

\[ + (\bar{Y}_{..} - \bar{Y}_{...}) + (\bar{Y}_{ij} - \bar{Y}_{...}) \right]^2 \]

\[ = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} 2 (\bar{Y}_{ij} - \bar{Y}_{..} - \bar{Y}_{ij} + \bar{Y}_{...}) (\bar{Y}_{ij} - \bar{Y}_{...}) \]

\[ = 2n \sum_{j=1}^{b} (\bar{Y}_{ij} - \bar{Y}_{..}) \frac{a}{n} \sum_{i=1}^{a} (\bar{Y}_{ij} - \bar{Y}_{..} - \bar{Y}_{ij} + \bar{Y}_{...}) \]

\[ = 0 \quad \text{because} \]

\[ \sum_{i=1}^{a} (\bar{Y}_{ij} - \bar{Y}_{..} - \bar{Y}_{ij} + \bar{Y}_{...}) = \]

\[ = a \bar{Y}_{ij} - a \bar{Y}_{..} - a \bar{Y}_{ij} + a \bar{Y}_{...} = 0 \]
Error Sum of Squares

$$\text{SSE} =$$

$$= \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (Y_{ijk} - \overline{Y}_{..})^2 = bn \sum_{i=1}^{a} (\overline{Y}_{i..} - \overline{Y}_{..})^2$$

It can be shown that

$$E[\text{SSA}] =$$
Thus
\[ MSA = \frac{SSA}{(a-1)} \] is an unbiased estimator of \( \sigma^2 + \frac{bn}{(a-1)} \sum_{i=1}^{a} \alpha_i^2 \)

Factor B Sum of Squares
\[ SSB = an \sum_{j=1}^{b} (\bar{Y}_{.j} - \bar{Y}_{..})^2 \]
and \[ MSB = \frac{SSB}{(b-1)} \] is an unbiased estimator

Interaction Sum of Squares
\[ SSAB = n \sum_{i=1}^{a} \sum_{j=1}^{b} (\bar{Y}_{ij} - \bar{Y}_{.j} - \bar{Y}_{i.} + \bar{Y}_{..})^2 \]
\[ E[SSAB] = (a-1)(b-1) \sigma^2 + n \sum_{i=1}^{a} \sum_{j=1}^{b} (\alpha \beta)_{ij}^2 \]
\[ MSAB = \frac{SSAB}{(a-1)(b-1)} \] is an unbiased estimator of
### Degrees of Freedom

- Total: $nab - 1$

### ANOVA Table

<table>
<thead>
<tr>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSA</td>
<td>$a-1$</td>
<td>MSA</td>
<td>$\frac{MSA}{MSE}$</td>
</tr>
<tr>
<td>SSB</td>
<td>$b-1$</td>
<td>MSB</td>
<td>$\frac{MSB}{MSE}$</td>
</tr>
<tr>
<td>SSAB</td>
<td>$(a-1)(b-1)$</td>
<td>MSAB</td>
<td>$\frac{MSAB}{MSE}$</td>
</tr>
<tr>
<td>SSE</td>
<td>$(n-1)ab$</td>
<td>MSE</td>
<td></td>
</tr>
<tr>
<td>SSTO</td>
<td>$nab-1$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The first step is always to test

$H_0: (\alpha \beta)_{ij} = 0$ for all $i,j$

vs

$H_a: \text{not } H_0$
If this test is significant, then testing the main effects is not needed, because we already have determined that we need to analyze the cell means $\mu_{ij}$ individually.

If the interaction test is not significant then we explore both main effects:

$H_0: \alpha_i = 0$ for all $i$ vs $H_a: \text{not } H_0$

Reject $H_0$ in favor of $H_a$ if

and

$H_0: \beta_j = 0$ for all $j$ vs $H_a: \text{not } H_0$

Reject $H_0$ in favor of $H_a$ if
Example 1: See SAS output.

When interactions are significant and meaningful, what do we do in follow-up? We analyze the individual cell means \( \bar{Y}_{ij} \) to determine which \( \mu_{ij} \)'s differ.

**Tukey**: Declare \( \mu_{ij} \neq \mu_{ij'} \) if

\[
\left| \bar{Y}_{ij} - \bar{Y}_{ij'} \right| \geq q(1-\alpha; ab, (n-1)ab) \sqrt{\frac{MSE}{n}}
\]

Example: \( q(.95; 9, 18) = 4.96 \)
Bonferroni: Declare \( \mu_{ij} \neq \mu_{ij'} \) if 

\[
|\bar{Y}_{ij} - \bar{Y}_{ij'}| \geq \sqrt{2 \text{MSE} \frac{t(1-\frac{\alpha}{2g};(n-1)ab)}{n}}
\]

where \( g \) would equal \( \binom{ab}{2} = \frac{ab(ab-1)}{2} \), if all pairs are being compared.

Example: \( \alpha = .05 \) \( \binom{9}{2} = 36 \)

Now suppose the interaction effect is not significant. Which main effects are significant (Both or Just One)?
If both main effects are significant then you want to know which levels among A levels differ and which B levels differ. That is, you want

The means \( \bar{Y}_{i.} \) and \( \bar{Y}_{.j} \) are not independent here. Why? So Tukey's method is not an option.

Bonferroni Declare \( \mu_i \neq \mu_j \) if
Example 2 - See SAS
\[ \alpha = 0.05 \quad n = 6 \quad a = 2 \quad b = 3 \]
\[ q = \binom{2}{2} + \binom{3}{2} = 4 \]
\[ t(1 - \frac{0.05}{2\times4}; 5(2)(3)) = 2.66 \]

Factor A \quad \bar{Y}_{1..} = 21.4 \quad \bar{Y}_{2..} = 26.1

Factor B \quad \bar{Y}_{1..} = 30.58 \quad \bar{Y}_{2..} = 24.08 \quad \bar{Y}_{3..} = 16.67

If the interactions are not significant and only one of the main effects is significant (suppose only B is),
then we want to analyze the means \( \bar{Y}_{1.}, \bar{Y}_{2.}, \ldots, \bar{Y}_{b.} \) in the follow-up step.

**Tukey** Why is it now appropriate?

Declare \( \mu_{ij} \neq \mu_{ij'} \) if

**Bonferroni**

Declare \( \mu_{ij} \neq \mu_{ij'} \) if

where \( g = \binom{b}{2} \).

The analysis scheme may be outlined in the following flowchart:
Two Factor - Factorial Design

Is overall ANOVA significant?

Yes

Are Interactions Significant

Yes

Compare all ab cell means to one another

No

Are

None

Stop Factors Not Useful

Both Are

Compare Main Effect Mean Levels to one another for Factor A

Only One Is

Compare Main Effect Mean Levels to one another for Factor B

Stop Factors Not Useful