Modeling Effects and Additive Two-Factor Models (i.e. without interaction)

STAT:5201

Week 5: Lecture 1
To “model” the data...

- ...to ‘break-down’ into its component parts.

- ...to define how the data were generated, or define the data generating mechanism.

- ...to show each observation as a sum of effects, each effect coming from a different source of variation.

- ....to specify a statistical distribution for the data.
Example (RCBD primary factor of \textit{daylength} and block on \textit{LITTER})

Hibernation example with 1 primary factor of \textit{daylength} and a block. For each of 6 random litters we have two hamsters, the treatments of \textit{short} and \textit{long} will be randomly assigned to each pair of litter hamsters.

The observed enzyme level is generated by taking the grand mean, adding a litter effect and a day length effect, plus adding a random effect representing unexplained variation, or ‘noise’.
There are two levels to the main effect of *daylength*, which we perceive as a bump-up or bump-down, depending on the given level (short or long).
It might be helpful to think of the residual as what is ‘left over’ after $\mu$ and the other effects have been removed.

Example (RCBD primary factor of **daylength** and block on **LITTER**)

In the formal mathematical model, we see the different effects present, and the distribution of all random effects (we consider block as random here).
In the previous RCBD example, we assumed no interaction between the block *LITTER* and *daylength*. This assumption is often made for a blocking factor & a primary factor.

A model that includes only main effects and no interaction is an additive model or a main-effects only model.

In the main-effects only model, the effects of one factor does not depend upon the level of the other factor. Changing from one level of factor A to another has the same effect across all levels of factor B.
When we estimate effects in our models, we will use a hat or “∧” to represent the estimate.

We will use the data to compute estimates of the effects. And the observed enzyme level can be written as a sum of these estimated effects.
We often consider the “fitted value” or the value predicted by the model:

\[ \hat{\text{Enzyme Level}} = \beta_0 + \hat{\beta}_1 \text{LITTER}_r + \hat{\beta}_2 \text{day length} \]

(Just doesn’t include the residual)

Continuing,

\[ \hat{\text{RESIDUAL}} = \hat{\text{ENZYME LEVEL}} - \text{Enzyme Level} \]

“observed response” - “fitted response”
Estimates for the Additive Model

- We will consider *litter* as a fixed effect in the model here.

Formal model (two fixed effects, no interaction):

\[ Y_{ij} = \mu + L_i + d_j + \epsilon_{ij} \text{ with } \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \]

\[ \text{litter } i = 1, \ldots, 6 \text{ and daylength } j = 1, 2 \text{ (short/long)} \]

**Estimates for the balanced scenario**, sum-to-zero constraints:

**Grand mean:** \( \hat{\mu} = \bar{Y}.. = \frac{\sum_i \sum_j Y_{ij}}{12} \)

**litter \( i \) effect:** \( \hat{L}_i = \text{litter mean} - \text{grand mean} = \bar{Y}_i. - \bar{Y}.. = \frac{\sum_{j=1}^{2} Y_{ij}}{2} - \bar{Y}.. \)

**daylength \( j \) effect:** \( \hat{d}_j = \text{daylength mean} - \text{grand mean} = \bar{Y}.j - \bar{Y}.. = \frac{\sum_{i=1}^{6} Y_{ij}}{6} - \bar{Y}.. \)

**Residual estimate:** \( \hat{\epsilon}_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - (\hat{\mu} + \hat{L}_i + \hat{d}_j) \)

We can think of \( \hat{\epsilon}_{ij} \) as the result of aligning the data for all of the effects in the model, including the overall mean.
### Estimates for the Additive Model

<table>
<thead>
<tr>
<th>Litter</th>
<th>S</th>
<th>L</th>
<th>( \hat{L}_i )</th>
<th>( \hat{d}_j )</th>
<th>( \hat{Y}_{ij} )</th>
<th>( \hat{E}_{ij} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.1</td>
<td>2.6</td>
<td>+0.275</td>
<td>-0.075</td>
<td>2.075</td>
<td>0.025</td>
</tr>
<tr>
<td>2</td>
<td>1.8</td>
<td>2.2</td>
<td>-0.075</td>
<td>+0.275</td>
<td>1.725</td>
<td>-0.075</td>
</tr>
<tr>
<td>3</td>
<td>1.4</td>
<td>2.4</td>
<td>-0.175</td>
<td>+0.275</td>
<td>1.625</td>
<td>0.075</td>
</tr>
<tr>
<td>4</td>
<td>1.2</td>
<td>1.7</td>
<td>-0.025</td>
<td>+0.275</td>
<td>1.175</td>
<td>-0.225</td>
</tr>
<tr>
<td>5</td>
<td>1.9</td>
<td>2.9</td>
<td>+0.325</td>
<td>-0.275</td>
<td>2.125</td>
<td>-0.225</td>
</tr>
<tr>
<td>6</td>
<td>2.4</td>
<td>2.3</td>
<td>+0.275</td>
<td>-0.275</td>
<td>2.075</td>
<td>0.325</td>
</tr>
</tbody>
</table>

Mean: 1.8, 2.35

\( \hat{\mu} = 2.075 \)

\( \bar{Y}_{.1} = 1.8 \)

\( \bar{Y}_{.2} = 2.35 \)

\( \hat{d}_i = \bar{Y}_{.i} - \bar{Y}_{..} \)

So, for \( Y_{11} \):

\[ Y_{11} = (\hat{\mu} + \hat{L}_1 + \hat{d}_1) + \hat{E}_{11} \]

\[ = 2.075 + 0.275 - 0.275 + 0.025 \]

\[ = 2.1 \]
Estimates for the Additive Model

Main effects Model (no interaction)

This model without interaction will provide a fit with parallel lines in the interaction plot.
See handout on fitting two-factor additive model in SAS.
Estimates for the Additive Model

Example (SAS two-factor additive model)

The other way to present the interaction plot.
Two-factor CRD with Interaction

Example (Animal-fattening experiment)

Two primary factors: Vitamin B12 (0mg, 5mg) & Antibiotics (0mg, 40mg)

Three animals randomly assigned to each of 4 treatments. Response was weight gain in lbs/week.

Because we can not assume that the two factors do not interact, we should include interaction in the model.

Two-way ANOVA with interaction will be the next model we consider.
A factorial model for data is called *hierarchical* if the presence of any term in the model implies the presence of all lower order terms.

As a general rule, if an interaction term is included in the model, then all of its component parts should also be included.

If you include this interaction: then, also include:

- $AB$  
- $ABC$

In our ANOVA tables, the lower-order terms will appear earlier (at the top), but we will start the testing of terms at the bottom of the table, for higher-order interactions first. If the interaction for $AB$ is significant, then we will not test for the main effects of $A$ and $B$ as those terms must remain in the model, regardless of significance.
Plausibility of the Model

- When we fit a model, we are probably using past experience or scientific knowledge to choose the model, and we’re making assumptions about the data.

- Before interpreting any results from the model fit, we should check to see if the model is plausible (reasonable).

  Usually, this is done by examining the residuals - the part of the variations that the model can’t explain.

- Things to consider:
  - Other possible effects that were not included
  - Interaction effects
  - Assumptions on the errors (e.g. $\epsilon_{ij} \sim \text{N}(0, \sigma^2)$)