ANOVA Decomposition (Cont’d)

Table 26: Wood Experiment : Summarized data for whole plot analysis

<table>
<thead>
<tr>
<th>Source</th>
<th>Rep 1</th>
<th>Rep 2</th>
<th>Rep 3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>181.1</td>
<td>224.7</td>
<td>219.0</td>
<td>624.8</td>
</tr>
<tr>
<td>$a_2$</td>
<td>168.0</td>
<td>191.0</td>
<td>128.8</td>
<td>487.8</td>
</tr>
<tr>
<td>Total</td>
<td>349.1</td>
<td>415.7</td>
<td>347.8</td>
<td>1112.6</td>
</tr>
</tbody>
</table>

The $SS_A$ can be calculated from the data in $LNP_{48}$.

\[
SS_A = \frac{(624.8^2 + 487.8^2) - 1112.6^2}{24} = 782.04.
\]

\[
SS_{Rep} = \frac{(349.1^2 + 415.7^2 + 347.8^2) - 1112.6^2}{24} = 376.99.
\]

\[
SS_{whole} = SS_{Rep \times A} = 398.37.
\]

\[
SS_{sub} = 927.88 - SS_{whole} - SS_{Rep} = 152.52.
\]

**Expected Mean Squares in ANOVA**

\[
\begin{align*}
W_0 \oplus T_i & \quad \text{(Replicate or block)} \\
S_i & \quad \text{dim}(S_i) = n_{R} \\
W_i & \quad \text{error} \\
S_2 & \quad \text{(Rep \times A)} \\
T_2 & \quad \text{Whole plot} \\
S_3 & \quad \text{(Rep \times B)} \times (Rep \times A \times B) \\
T_3 & \quad \text{Subplot} \\
\end{align*}
\]

- Proofs are similar but more tedious than in one-way random effects model ($LNP_{3,33-36}$).
Hypothesis Testing

\[ F_A = \frac{MS_A}{MS_{\text{whole}}} \Rightarrow H_0^1: \alpha_1 = \cdots = \alpha_I, \]

Under \( H_0^2 \), \( \Delta \text{SS}_A \sim e_2^2 \chi_{2(d_m(w))}^2 \)

\[ F_B = \frac{MS_B}{MS_{\text{sub}}} \Rightarrow H_0^2: \beta_1 = \cdots = \beta_I, \]

Under \( H_0^2 \), \( \Delta \text{SS}_B \sim e_2^2 \chi_{2(d_m(w))}^2 \)

Apply similar argument as for \( H_0^2 \)

\[ F_{\Delta B} = \frac{MS_{\Delta B}}{MS_{\text{sub}}} \Rightarrow H_0^3: (\alpha \beta)_{ij} = \text{constant}, \]

\[ F_{\text{Rep}} = \frac{MS_{\text{Rep}}}{MS_{\text{whole}}} \Rightarrow H_0^4: \sigma_R = 0. \]

Q: Why does \( A \) become insignificant?

Q: Why does \( B \) become significant?

Correct ANOVA Analysis

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Sum of Squares</th>
<th>Mean Squares</th>
<th>( F )</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S_1 )</td>
<td>3-1 = 2</td>
<td>376.99</td>
<td>188.50</td>
<td>0.95</td>
<td>0.513</td>
</tr>
<tr>
<td>( W_i )</td>
<td>2-1 = 1</td>
<td>782.04</td>
<td>782.04</td>
<td>3.93</td>
<td>0.186</td>
</tr>
<tr>
<td>( T_2 )</td>
<td>3-1 = 2</td>
<td>398.37</td>
<td>199.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( W_2 )</td>
<td>4-1 = 3</td>
<td>266.00</td>
<td>88.67</td>
<td>6.98</td>
<td>0.006</td>
</tr>
<tr>
<td>( W_3 )</td>
<td>(4-1)(5-1) = 12</td>
<td>62.79</td>
<td>20.93</td>
<td>1.65</td>
<td>0.230</td>
</tr>
<tr>
<td>( S_3 )</td>
<td>18-3-3 = 12</td>
<td>152.52</td>
<td>12.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \Omega )</td>
<td>23</td>
<td>2038.72</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3 \( (n_w) \) blocks, size 2
6 \( (n_R) \) whole plots, size 4
24 \( (N) \) sub-plots

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicate</td>
<td>2</td>
<td>376.99</td>
<td>188.50</td>
<td>0.95</td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>782.04</td>
<td>782.04</td>
<td>3.93</td>
</tr>
<tr>
<td>Whole plot error</td>
<td>2</td>
<td>398.37</td>
<td>199.19</td>
<td></td>
</tr>
<tr>
<td>( B \times B )</td>
<td>3</td>
<td>266.00</td>
<td>88.67</td>
<td>6.98</td>
</tr>
<tr>
<td>Subplot error</td>
<td>12</td>
<td>152.52</td>
<td>12.71</td>
<td></td>
</tr>
</tbody>
</table>

Subtotal: 23 | 2038.72

Randomized Block Design

**Analysis Results**

- Only \( B \) is significant.
- Explanation for discrepancy:

\[
MS_{\text{whole}} = 199.19 \gg MS_{\text{Residual}} = 57.99 \gg MS_{\text{sub}} = 12.71.
\]

- To test \( H_0^B : \sigma_B = 0 \), use

\[
F_{2, 2} \overset{\text{null dist.}}{\sim} \frac{MS_{\text{Rep}}}{MS_{\text{whole}}} = \frac{188.5}{199.19} = 0.95. < 1 \Rightarrow \hat{\sigma}_B^2 = 0
\]

⇒ no significant difference between three replications.

- When does testing \( H_0^B \) make sense?

**Reading:** textbook, 3.9
A Brief Note on Strip-Plot Design
- Major distinction between strip-plot design and split-plot design
  - strip-plot design: crossing (EU) plot structure; nesting (SP nested in WP)
  - split-plot design: random factor
- An example: laundry experiment
  - A: 2 washing machines \( a_1, a_2 \); B: 3 dryers \( b_1, b_2, b_3 \); EUs: cloth samples
    - Completely randomized design
      - do washing \( k_l \) times
      - do drying \( k_l \) times
    - Split-plot design
      - A: WP factor, B: SP factor
      - do washing \( k \) times
      - do drying \( l \) times
    - Strip-plot design
      - A: row plot factor
      - B: column plot factor
      - do washing \( k \) times
      - do drying \( l \) times

Further Reading:
Cheng (2014), Theory of Factorial Design, Chapter 1, Sec. 7.11

Transformation of Response
- Transform \( y \) before fitting a regression model.
- Theory: Suppose in the model
  \[
  y_i = \mu + \varepsilon_i, \quad \sigma^2 = \frac{\text{Var}(y_i)}{\text{Var}(\varepsilon_i)} = \frac{\sigma^2}{\sigma^2} = \frac{1}{\sigma^2}
  \]
  This can be detected by plotting residuals
  \[
  r_{ij} = y_{ij} - \bar{y}_i \quad \text{(for replicated experiment)} \quad \text{or} \quad r_i = y_i - \hat{y}_i \quad \text{(for unreplicated experiment)}
  \]
  (What pattern to look for?)
- Error transmission formula:
  \[
  z_i = f(y_i) \approx f(\mu) + f'(\mu)(y_i - \mu)
  \]
  by (\#)
  \[
  \sigma^2 = \frac{\text{Var}(z_i)}{\text{Var}(y_i)} \approx \left( f'(\mu) \right)^2 \sigma^2 = \left( f'(\mu) \right)^2 \sigma^2
  \]
  \[
  f(u) = \int f'(u) \, du \propto \int u^{-\alpha} \, du = \left\{ \begin{array}{ll}
  u^{1-\alpha}, & \text{if } \alpha \neq 1, \\
  \frac{\ln(u)}{\ln(\alpha)}, & \text{if } \alpha = 1
  \end{array} \right.
  \]
  for \( \alpha > 0 \)
**Power (Box-Cox) Transformation**

Model (✓):

\[ f_X(y) = \frac{y^\lambda - 1}{\lambda}, \quad \lambda \neq 0, \]

\[ \ln y_X, \quad \lambda = 0, \]

\[ z_X = f_X(y_X) = \begin{cases} 
\frac{y^\lambda - 1}{\lambda}, & \lambda \neq 0, \\
\ln y_X, & \lambda = 0.
\end{cases} \]

\[ f'_X(\mu_X) = \frac{\mu^\lambda - 1}{\lambda}, \]

\[ \sqrt{\text{Var}(z_X)} \approx |f'_X(\mu_X)| \sigma_X = \frac{\mu^\lambda - 1}{\lambda} \sigma_X \propto \mu^\lambda \mu_X = \mu^{\lambda + 1}. \]

- Choosing \( \lambda = 1 - \alpha \) would make \( \text{Var}(z_X) \) nearly constant over \( x \).
- Since \( \alpha \) is unknown, \( \lambda \) can be chosen by some statistical criterion (e.g., maximum likelihood). A simpler method is to try a few selected values of \( \lambda \) (see Table 28 in LNP.4-70). In each transform, analyze the \( z_X \) data and choose the transformation (i.e., \( \lambda \) value) such that
  - (a) it gives a parsimonious model, "few" very significant effects
  - (b) no unusual pattern in the residual plots,
  - (c) good interpretation of the transformation.

Example of (c):

\[ \bar{y}_X = \text{survival time}, \quad \bar{y}_X^{-1} = \text{rate of dying} \]

in the example of Box-Cox(1964).

---

**Variance Stabilizing Transformations**

Their relationship may be identified from:

(i) residual plot of \( \hat{e}_i \) vs. \( \hat{y}_i \) \( \Rightarrow \alpha (= 1 - \lambda) \)

(ii) MLE (or confidence interval) of \( \lambda \)

Table 28: Variance Stabilizing Transformations

<table>
<thead>
<tr>
<th>( \sigma_X \propto \mu_X^\alpha )</th>
<th>( \alpha )</th>
<th>( \lambda (= 1 - \alpha) )</th>
<th>Transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma_X \propto \mu_X^3 )</td>
<td>3</td>
<td>-2</td>
<td>reciprocal squared</td>
</tr>
<tr>
<td>( \sigma_X \propto \mu_X^2 )</td>
<td>2</td>
<td>-1</td>
<td>reciprocal</td>
</tr>
<tr>
<td>( \sigma_X \propto \mu_X^{3/2} )</td>
<td>3/2</td>
<td>-1/2</td>
<td>reciprocal square root</td>
</tr>
<tr>
<td>( \sigma_X \propto \mu_X )</td>
<td>1</td>
<td>0</td>
<td>log</td>
</tr>
<tr>
<td>( \sigma_X \propto \mu_X^{1/2} )</td>
<td>1/2</td>
<td>1/2</td>
<td>square root</td>
</tr>
<tr>
<td>( \sigma_X \propto \text{constant} )</td>
<td>0</td>
<td>1</td>
<td>original scale</td>
</tr>
<tr>
<td>( \sigma_X \propto \mu_X^{-1/2} )</td>
<td>-1/2</td>
<td>3/2</td>
<td>3/2 power</td>
</tr>
<tr>
<td>( \sigma_X \propto \mu_X^{-1} )</td>
<td>-1</td>
<td>2</td>
<td>square</td>
</tr>
</tbody>
</table>

- no transformation

- transformations with good interpretation
Analysis of Drill Experiment

- Data in Table 3.40 of textbook (p.135).
  - four factors $A$, $B$, $C$ and $D$, each at two levels
  - use a $2^4$ design $\rightarrow$ full factorial design
  - fit a model with 4 main effects and 6 two-factor interactions ($2^4$'s)

The $t$-vs-$\hat{y}$ plot shows an increasing pattern.

![Graph showing the pattern between $\sigma$ and $\mu$](image)

$\sigma_{\hat{y}} \propto \mu_{\hat{y}}$

$\Rightarrow \alpha = 1$

$\Rightarrow \lambda = 1 - \alpha = 0$

$\Rightarrow$ suggest log-transformation of $y_{\hat{y}}$

Figure 2: $r_I$ vs. $\hat{y}_I$, Drill Experiment

Scaled lambda plot

- For each of the eight transformations $\lambda$ values in Table 28 (Lnp.4-70), a model of main effects and $2^4$'s is fitted to the transformed $z_{\hat{y}} = f_\lambda (y_{\hat{y}})$. The $t$-statistic values for the 10 effects are displayed in Figure 3 (Lnp.4-73).

- Comments on the plot.
  - For the log transformation ($\lambda = 0$), the largest $t$ statistics ($C$, $B$, and $D$) stand out.
  - The next best is $\lambda = -1/2$, but not as good as log transformation (Why? It has a $2^4 BC$, but the log transform removes the term $BC$.)
  - On the original scale ($\lambda = 1$), the four main effects do not separate apart.

- Conclusion: Use log transformation.

Q: Why not draw $\hat{\beta}_\lambda$?

Q: Why use $t$-statistic?

1. Eliminate unit
   (Note, $z_{\hat{y}}$ has different units for different $\lambda$'s)

2. For different $\lambda$'s, can use same critical value to declare significance
Scaled lambda plot: Drill Experiment

(exercise) use the Box-Cox transformation method to obtain the MLE and confidence interval of \( \lambda \).

**Note 1.** Observe how effect significance changes with \( \lambda \).

**Note 2.** The plot does not show how good the fitting is (say, \( R^2 = ? \)).

![Figure 3: Scaled \( \lambda \) Plot](image)

- can use a simpler model to explain the response

**Reading:** textbook, 3.11