Two-Factor ANOVA model with \( n = 1 \) (no replication)

1. Two-factor ANOVA model with \( n = 1 \) (no replication)
   1.1 Two-factor model without interaction
       - Sum of squares
       - ANOVA Table
       - \( F \) tests (for main effects)
       - Estimation of means
   1.2 Example: Insurance
       - Sum of squares:
   1.3 Checking for the presence of interaction: Tukey's test for additivity
       - Estimation of \( D \)

Contributors

1. Two-factor ANOVA model with \( n = 1 \) (no replication)

- For some studies, there is only one replicate per treatment, i.e., \( n = 1 \).
- ANOVA model for two-factor studies need to be modified, since
  - the degrees of freedom associated with \( \text{SSE} \) will be \( \left( (n - 1)ab = 0 \right) \);
  - thus the error variance \( \text{Variance (SSE)} \) can not be estimated by \( \text{Variance (SSE)} \) anymore.

- Idea: make the model simpler by assuming the two factors do not interact with each other. Validity of this assumption needs to be checked.
1.1 Two-factor model without interaction

With \( n = 1 \).

- Model equation:
  \[
  Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}, \quad i = 1, \ldots, a, \quad j = 1, \ldots, b.
  \]

- Identifiability constraints:
  \[
  \sum_{i=1}^{a} \alpha_i = 0, \quad \sum_{j=1}^{b} \beta_j = 0.
  \]

- Distributional assumptions: \( \epsilon_{ij} \) are i.i.d. \( N(0, \sigma^2) \)

**Sum of squares**

Interaction sum of squares now plays the role of error sum of squares.

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

\[
MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)} \text{ since } d.f.(SS_{AB}) = (a-1)(b-1).
\]

- In the general two-factor ANOVA model (when \( n = 1 \)),
  \[
  E(MS_{AB}) = \sigma^2 + \frac{\sum_{i=1}^{a} \sum_{j=1}^{b} (\alpha \beta)_{ij}^2}{(a-1)(b-1)}
  \]

  - Under the model without interaction: \( E(MS_{AB}) = \sigma^2 \)
  - Thus \( MS_{AB} \) can be used to estimate \( \sigma^2 \).

**ANOVA Table**

ANOVA table for two-factor model without interaction and \( (n=1) \)

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor A</td>
<td>( SSA = b \sum_i (\overline{Y}<em>{i.} - \overline{Y}</em>{..})^2 )</td>
<td>((a - 1))</td>
<td>(MSA)</td>
</tr>
<tr>
<td>Factor B</td>
<td>( SSB = a \sum_j (\overline{Y}<em>{.j} - \overline{Y}</em>{..})^2 )</td>
<td>((b - 1))</td>
<td>(MSB)</td>
</tr>
<tr>
<td>Error</td>
<td>( SS_{AB} = \sum_{i=1}^{a} \sum_{j=1}^{b} (\overline{Y}<em>{ij} - \overline{Y}</em>{i.} - \overline{Y}<em>{.j} + \overline{Y}</em>{..})^2 )</td>
<td>((a - 1)(b - 1))</td>
<td>(MS_{AB})</td>
</tr>
</tbody>
</table>
\[
\text{Total} \quad \frac{(\overline{Y}_{..})^2}{\text{ab} - 1}
\]

Expected mean squares (under no interaction):

\[
\begin{align*}
E(MSA) &= \sigma^2 + \frac{b\sum_{i=1}^{a}\alpha_i^2}{a - 1}, \quad E(MSB) = \sigma^2 + \frac{a\sum_{j=1}^{b}\beta_j^2}{b - 1}, \quad E(MSAB) = \sigma^2
\end{align*}
\]

\[F\text{ tests (for main effects)}\]

Test factor A main effects: \(H_o: \alpha_1 = ... = \alpha_a = 0\) vs. \(H_a:\) not all \(\alpha_i\)'s are equal to zero.

- \(F_A^* = \frac{MSA}{MSAB} \sim F_{a - 1, (a - 1)(b - 1)}\) under \(H_o\).
- Reject \(H_o\) at level of significance \(\alpha\) if observed \(F_A^* > F(1 - \alpha; a - 1, (a - 1)(b - 1))\).

Test factor B main effects: \(H_o: \beta_1 = ... = \beta_b = 0\) vs. \(H_a:\) not all \(\beta_j\)'s are equal to zero.

- \(F_B^* = \frac{MSB}{MSAB} \sim F_{b - 1, (a - 1)(b - 1)}\) under \(H_o\).
- Reject \(H_o\) at level of significance \(\alpha\) if observed \(F_B^* > F(1 - \alpha; b - 1, (a - 1)(b - 1))\).

\[\text{Estimation of means}\]

Estimation of factor level means \(\mu_{(i.j)}\)'s , \(\mu_{(i.j)}\)'s.

- Proceed as before, viz., use the unbiased estimator \(\overline{Y}_{(i.j)}\) for \(\mu_{(i.j)}\) and \(\overline{Y}_{(i.j)}\) for \(\mu_{(i.j)}\), but replace \(\text{MSE}\) by \(\text{MSAB}\) and use the degrees of freedom of \(\text{MSAB}\), that is \((a - 1)(b - 1))\). Thus, estimated standard errors:

\[
\begin{align*}
\text{s}(\overline{Y}_{(i.j)}) &= \sqrt{\text{MSAB} \left(\frac{1}{b} + \frac{1}{a} - \frac{1}{ab}\right)} = \sqrt{\text{MSAB} \left(\frac{a + b - 1}{ab}\right)}
\end{align*}
\]

1.2 Example: Insurance

An analyst studied the premium for auto insurance charged by an insurance company in six cities. The six cities were selected to represent different sizes (Factor A: small, medium, large) and different regions of the state (Factor B: east, west). There is only one city for each combination of size and region. The amounts of premiums charged for a specific type of coverage in a given risk category for each of the six cities are given in the following table.


Updated: Tue, 08 Jun 2021 22:20:19 GMT

Powered by
Table 1: Numbers in parentheses are \( \widehat{\mu}_{ij} = \overline{Y}_{i.} + \overline{Y}_{.j} - \overline{Y}_{..} \)

<table>
<thead>
<tr>
<th>Factor A</th>
<th>Factor B</th>
<th>East</th>
<th>West</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small 140(135)</td>
<td>100(105)</td>
<td>100(105)</td>
<td>( \overline{Y}_{1.} = 120 )</td>
</tr>
<tr>
<td>Medium 210(210)</td>
<td>180(180)</td>
<td>( \overline{Y}_{2.} = 195 )</td>
<td></td>
</tr>
<tr>
<td>Large 220(225)</td>
<td>200(195)</td>
<td>( \overline{Y}_{3.} = 210 )</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: ANOVA Table for Insurance example

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor A</td>
<td>( SSA = 9300 )</td>
<td>( a - 1 = 2 )</td>
<td>( MSA = 4650 )</td>
</tr>
<tr>
<td>Factor B</td>
<td>( SSB = 1350 )</td>
<td>( b - 1 = 1 )</td>
<td>( MSB = 1350 )</td>
</tr>
<tr>
<td>Error</td>
<td>( SSAB = 100 )</td>
<td>( (a - 1)(b - 1) = 2 )</td>
<td>( MSAB = 50 )</td>
</tr>
<tr>
<td>Total</td>
<td>( SSTO = 10750 )</td>
<td>( ab - 1 = 5 )</td>
<td></td>
</tr>
</tbody>
</table>

\( F_A^* = \frac{MSA}{MSAB} = \frac{4650}{50} = 93 \) and \( F(0.95; 2, 2) = 19 \). Thus reject \( H_0 \) at level 0.05.

Interaction plot based on the treatment sample means \( \{Y_{ij}\} \)'s: no strong interactions.

**Sum of squares:**

- Here \( (a = 3), (b = 2), (n = 1) \).
- \( SSA = 2[(120 - 175)^2 + (195 - 175)^2 + (210 - 175)^2] = 9300 \).
- \( SSB = 3[(190 - 175)^2 + (160 - 175)^2] = 1350 \).
- \( SSAB = (140 - 120 - 190 + 175)^2 + ... + (200 - 210 - 160 + 175)^2 = 100 \).
- \( SSTO = SSA + SSB + SSAB = 10750 \).

Hypothesis testing:
- Test \( H_0: \mu_{1.} = \mu_{2.} = \mu_{3.} \) (equivalently, \( H_0: \alpha_1 = \alpha_2 = \alpha_3 = 0 \)) at level 0.05.

\( F_A^* = \frac{MSA}{MSAB} = \frac{4650}{50} = 93 \) and \( F(0.95; 2, 2) = 19 \). Thus reject \( H_0 \) at level 0.05.

- Estimation of \( \{\mu_{i.}\} \): e.g.,
  \( \{\mu_{1.}\} = \overline{Y}_{1.} + \overline{Y}_{.1} - \overline{Y}_{..} = 120 + 190 - 175 = 135 \).
- Estimation of \( \{\mu_{.j}\} \) and \( \{\mu_{ij}\} \): e.g.,
  \( \{\mu_{1.}\} = \overline{Y}_{1.} = 120 \).
  \( s(\overline{Y}_{1.}) = \sqrt{\frac{MSAB}{b}} = \sqrt{\frac{50}{2}} = 5 \).
  The 95% C.I. for \( \mu_{1.} \) is:
  \( \overline{Y}_{1.} \pm t(0.975; 2) * s(\overline{Y}_{1.}) = 120 \pm 4.3*5 = (98.5, 141.5) \).
1.3 Checking for the presence of interaction: Tukey’s test for additivity

For a two-factor study with \(n = 1\), decide whether or not the two factors are interacting.

- In the no-interaction model, we assume that all \(\langle \alpha \beta \rangle_{ij} = 0\).
- Idea: use a less severe restriction on the interaction effects, by assuming
  \(\langle \alpha \beta \rangle_{ij} = D \alpha_i \beta_j, i = 1, \ldots, a, j = 1, \ldots, b, \rangle\)
  where \(D\) is an unknown parameter.
- The model becomes:
  \(Y_{ij} = \mu_{..} + \alpha_i + \beta_j + D \alpha_i \beta_j + \epsilon_{ij}, i = 1, \ldots, a, j = 1, \ldots, b, \rangle\)
  under the constraints that
  \(\sum_{i = 1}^{a}\alpha_i = \sum_{j = 1}^{b}\beta_j = 0.\)

**Estimation of \(D\)**

- Multiply \(\langle \alpha \beta \rangle_{ij}\) on both sides of the equation:
  \(\langle \alpha \beta \rangle_{ij}Y_{ij} = \mu_{..} \alpha_i \beta_j + \alpha_i^2 \beta_j + \alpha_i \beta_j^2 + D \alpha_i^2 \beta_j^2 + \epsilon_{ij} \alpha_i \beta_j\)
- Sum over all pairs (i, j):
  \(\sum_{i=1}^{a}\sum_{j=1}^{b}\alpha_i \beta_jY_{ij} = D \sum_{i=1}^{a}\sum_{j=1}^{b}\alpha_i^2 \beta_j^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\epsilon_{ij} \alpha_i \beta_j\)
- Then
  \(\widetilde{D} := \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\alpha_i \beta_jY_{ij}}{\sum_{i=1}^{a}\sum_{j=1}^{b}\alpha_i^2 \beta_j^2} \approx D\)
- We have the following estimates:
  \(\hat{\alpha}_i = \overline{Y}_{i.} - \overline{Y}_{..}, \hat{\beta}_j = \overline{Y}_{.j} - \overline{Y}_{..}\)
  Thus, an estimator of \(D\) (which is also the least squares and the maximum likelihood estimator) is given by
  \(\hat{D} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{Y}_{i.} - \overline{Y}_{..})(\overline{Y}_{.j} - \overline{Y}_{..})Y_{ij}}{\sum_{i=1}^{a}(\overline{Y}_{i.} - \overline{Y}_{..})^2(\sum_{j=1}^{b}(\overline{Y}_{.j} - \overline{Y}_{..})^2)}\)

**ANOVA decomposition**

\[\text{SSTO} = \text{SSA} + \text{SSB} + \text{SSAB} + \text{SSRem}.\]

- Interaction sum of squares
  \(\text{SSAB} = \sum_{i=1}^{a}\sum_{j=1}^{b}\hat{\alpha}_i^2 \hat{\beta}_j^2 = \frac{1}{a \cdot b} \left( \sum_{i=1}^{a}(\overline{Y}_{i.} - \overline{Y}_{..})^2(\overline{Y}_{.j} - \overline{Y}_{..})^2 \right) \)
- Remainder sum of squares
  \(\text{SSRem} = \text{SSTO} - \text{SSA} - \text{SSB} - \text{SSAB}\)
- Decomposition of degrees of freedom
  \(\text{df}(\text{SSTO}) = \text{df}(\text{SSA}) + \text{df}(\text{SSB}) + \text{df}(\text{SSAB}) + \text{df}(\text{SSRem})\)
  \(\text{df} = (a - 1) + (b - 1) + 1 + (a - 1)(b - 1)\)
- Tukey’s one degree of freedom test for additivity: \(H_0: D = 0\) (i.e., no interaction) vs. \(H_a: D \neq 0\).
  \(F\) ratio \(F_{\text{Tukey}}^{*} = \frac{\text{SSAB}^*/1}{\text{SSRem}^*/(ab - a - b)} \sim F_{1, ab - a - b}\)
• Decision rule: reject \( \{H_o: D = 0\} \) at level of significance \( \{\alpha\} \) if \( \{F_{(\text{Tukey})}^*\} > F(1 - \{\alpha\}; 1, ab - a - b) \).

Example: Insurance

• \( \sum_{ij}(\overline{Y}_{i.} - \overline{Y}_{..})(\overline{Y}_{.j} - \overline{Y}_{..})Y_{ij} = -13500. \)

• \( \sum_{i=1}^{a}(\overline{Y}_{i.} - \overline{Y}_{..})^2 = 4650 \), and \( \sum_{j=1}^{b}(\overline{Y}_{.j} - \overline{Y}_{..})^2 = 450. \)

• \( SS_{AB}^* = \frac{(-13500)^2}{4650 * 450} = 87.1. \)

• \( SS_{Rem}^* = 10750 - 9300 - 1350 - 87.1 = 12.9. \)

• \( ab - a - b = 3*2 - 3 - 2 = 1. \)

• \( \{F\}-\text{ratio for Tukey's test:} \)
  \[ F_{(\text{Tukey})}^* = \frac{SS_{AB}^*}{1} = \frac{SS_{Rem}^*}{1} = \frac{87.1}{12.9} = 6.8. \]

• When \( \{\alpha\} = 0.05, F(0.95; 1, 1) = 161.4 > 6.8. \)

• Thus, we can not reject \( \{H_o: D = 0\} \) at the 0.05 level, and we conclude that there is no significant interaction between the two factors.

• Indeed, the p-value is \( \{p = P(F_{1,1} > 6.8) = 0.23\} \) which is not at all significant.

Contributors

• Scott Brunstein (UCD)
• Debashis Paul (UCD)