ANOVA table inference

ANOVA – ANalysis Of VAriance: provides an alternative approach for inferences.

- Test $H_0: \beta_1 = 0$.

- Compare two models:
  1. $y_i = \mu + e_i \rightarrow \hat{y}_i = \bar{y}$.
  2. $y_i = \beta_0 + \beta_1 x_i + e_i \rightarrow \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$.

- Measure of lack of fit using sums of squares.
  1. $SS_1 = \sum (y_i - \bar{y})^2 = SS_{total}$
  2. $SS_2 = \sum (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2 = SS_{error} = SS_{residual}$

- Regression model fits well when

  $$SS_{error} < SS_{total}$$ significantly
- \( SS_{model} = SS_{regression} = SS_{total} - SS_{error} \)

- Collect sums of squares into an ANOVA table:

<table>
<thead>
<tr>
<th>source of variation</th>
<th>df</th>
<th>sums of squares</th>
<th>mean square</th>
</tr>
</thead>
<tbody>
<tr>
<td>model</td>
<td>1</td>
<td>( SS_{model} )</td>
<td>( SS_{model}/1 )</td>
</tr>
<tr>
<td>error</td>
<td>n-2</td>
<td>( SS_{error} )</td>
<td>( SS_{error}/(n - 2) )</td>
</tr>
<tr>
<td>total</td>
<td>n-1</td>
<td>( SS_{total} )</td>
<td></td>
</tr>
</tbody>
</table>

- Note that \( MS_{error} \) is the variance estimate \( S_e^2 \).

- \( R^2 = SS_{model}/SS_{total} \) is the amount of variation in the response that is explained by the model.
ANOVA Table Decomposition

\[ SS_{total} = \sum_i (y_i - \bar{y})^2 \]

\[ = \sum_i (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 \]

\[ = \sum_i (y_i - \hat{y}_i)^2 + \sum_i (\hat{y}_i - \bar{y})^2 + 2 \sum_i (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) \]

\[ = \sum_i (y_i - \hat{y}_i)^2 + \sum_i (\hat{y} - \bar{y})^2 \]

\[ = SS_{error} + SS_{model} \]

Total variation = Variation explained by model + Variation in error
Inference by ANOVA table

- \( F = \frac{MS_{\text{model}}}{MSE} \) will have an \( F_{1,n-2} \) distribution if \( \beta_1 = 0 \) and tend to be large otherwise.

- Reject \( H_0 : \beta_1 = 0 \) vs \( H_1 : \beta_1 \neq 0 \) if

\[
F = \frac{MS_{\text{model}}}{MSE} > F_{(1,n-2),1-\alpha}
\]
Forbes Data Example

Total n=17 observations.

<table>
<thead>
<tr>
<th>source of variation</th>
<th>df</th>
<th>sums of squares</th>
<th>mean square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>model</td>
<td>1</td>
<td>0.22573</td>
<td><strong>MS_{model}=0.22573</strong></td>
<td>2961.55</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>error</td>
<td>15</td>
<td>0.00114</td>
<td><strong>MSE=0.00007622</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>16</td>
<td>0.22688</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Read Textbook Chapter 7 for ANOVA table.