Analysis of Variance (ANOVA)

W. Robert Stephenson
Department of Statistics
Iowa State University

One of the hallmarks of statistical thinking is the importance of measuring and understanding variability. The Shewhart Control Chart, which separates special cause from common cause variation, is one of the most important tools for understanding the current state of a process. The analysis of variance (ANOVA) is another statistical tool for splitting variability into component sources. These components can be thought of as the signal and the noise. The signal is seen as differences among group means. The noise is seen as variability within groups. By measuring the variability within groups one has a baseline against which differences among group means can be compared.

Example: The data below are from an experiment which studied how absorption of moisture in concrete is affected by different aggregates. Five different concrete aggregates were studied by exposing six samples of each to moisture for 2 days. The response variable is percent moisture absorbed.

<table>
<thead>
<tr>
<th>Aggregate</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>563</td>
<td>417</td>
<td>639</td>
<td>595</td>
<td>551</td>
</tr>
<tr>
<td></td>
<td>631</td>
<td>449</td>
<td>615</td>
<td>580</td>
<td>457</td>
</tr>
<tr>
<td></td>
<td>524</td>
<td>517</td>
<td>511</td>
<td>478</td>
<td>450</td>
</tr>
<tr>
<td></td>
<td>613</td>
<td>438</td>
<td>570</td>
<td>583</td>
<td>601</td>
</tr>
<tr>
<td></td>
<td>656</td>
<td>415</td>
<td>648</td>
<td>631</td>
<td>493</td>
</tr>
<tr>
<td></td>
<td>679</td>
<td>554</td>
<td>677</td>
<td>517</td>
<td>532</td>
</tr>
</tbody>
</table>

\[
\bar{Y}_i = \begin{array}{c}
611 \\
465 \\
610 \\
564 \\
514 \\
\end{array}
\]

\[
s_i^2 = \begin{array}{c}
3.385.2 \\
3.282.8 \\
3.640.0 \\
3.134.4 \\
3.409.6 \\
\end{array}
\]

\[
s_i = \begin{array}{c}
58.2 \\
57.3 \\
60.3 \\
56.0 \\
58.4 \\
\end{array}
\]

The idea of the analysis of variance is to take a summary of the variability in all the observations and partition it into separate sources. The summary is the sum of squares total, \( SS_{Total} \).

\[
SS_{Total} = \sum (Y - \bar{Y})^2 = \sum Y^2 - \frac{(\sum Y)^2}{N}
\]

where the summation is over all, \( N \), observations and \( \bar{Y} \) is the grand sample average. Alternatively,

\[
SS_{Total} = (N - 1)s^2
\]

where \( s^2 \) is the sample variance for the entire set of \( N \) observations.
Example:

\[ N = 30 \quad \sum Y = 16,584 \quad \bar{Y} = 552.8 \quad \sum Y^2 = 9,347,888 \]

\[ SS_{Total} = \left[ 9,347,888 - \frac{(16,584)^2}{30} \right] = 180,252.8 \]

This \textbf{sum of squares total} is partitioned into two separate, and additive, pieces. These are a \textbf{sum of squares among}, \( SS_{Among} \) and a \textbf{sum of squares within}, \( SS_{Within} \). The \( SS_{Within} \) accumulates variability from within each group.

\[ SS_{Within} = \sum (n_i - 1)s_i^2 \]

where \( s_i^2 \) is the sample variance of the \( i^{th} \) group and \( n_i \) is the number of observations in the \( i^{th} \) group.

Example:

\[ SS_{Within} = 5(3,384.2) + 5(3,282.8) + 5(3,640.0) + 5(3,134.4) + 5(3409.6) = 84,260 \]

The \textbf{sum of squares among}, \( SS_{Among} \) measures variability due to differences among the group means.

\[ SS_{Among} = \sum n_i(\bar{Y}_i - \bar{Y})^2 \]

where \( \bar{Y}_i \) is the sample mean for the \( i^{th} \) group.

Example:

\[
SS_{Among} = 6(611 - 552.8)^2 + 6(465 - 552.8)^2 + 6(610 - 552.8)^2 \\
+ 6(564 - 552.8)^2 + 6(514 - 552.8)^2 \\
= 6(58.2)^2 + 6(-87.8)^2 + 6(57.2)^2 + 6(11.2)^2 + 6(-38.8)^2 \\
SS_{Among} = 95,992.8
\]

Note that \( SS_{Among} + SS_{Within} = SS_{Total} \).

Associated with each sum of squares is a \textbf{degrees of freedom}. In general, one starts with \( N \) degrees of freedom and loses one degree of freedom for every sample mean calculated. For the \( SS_{Total} \) there is one grand sample average, therefore there are \( N - 1 \) degrees of freedom. There are \( n_i - 1 \) degrees of freedom within each group. Therefore, there are \( \sum(n_i - 1) = N - k \) degrees of freedom for \( SS_{Within} \). That leaves \( k - 1 \) degrees of freedom for \( SS_{Among} \).
A sum of squares divided by its associated degrees of freedom produces a **mean square**. The sums of squares, degrees of freedom, and mean squares are all summarized in an analysis of variance (ANOVA) table.

**Example:**

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among</td>
<td>95,992.8</td>
<td>4</td>
<td>23,998.2</td>
</tr>
<tr>
<td>Within</td>
<td>84,260.0</td>
<td>25</td>
<td>3,370.4</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>180,252.8</strong></td>
<td><strong>29</strong></td>
<td></td>
</tr>
</tbody>
</table>

Each mean square captures different elements of variability. The mean square within is often referred to as the **mean square error**. Within sample variability is attributed to random (sampling) error. This is the baseline against which differences among group means are compared. The mean square among contains some of this error variability but also variability due to differences among group means. The ratio $F = \frac{MS_{Among}}{MS_{Within}}$ serves as a measure of the statistical importance or **significance** of the differences among the group means. Values of $F$ close to one indicate that the differences among group means can be attributed to natural or random error variability. Values of $F$ much larger than one indicate that some of the groups differ significantly in terms of their mean or average values. The cutoff between “close to one” and “much larger than one” can be found in a table of the $F$ distribution. This tabulation assumes that the original observations are normally distributed with a common error variance.

**Example:**

$$F = \frac{23,998.2}{3,370.4} = 7.12$$

A critical value of $F$ with 4 degrees of freedom for the numerator and 25 degrees of freedom for the denominator is 2.76. Since 7.12 is greater than this critical value, one concludes that some of the concrete aggregates differ significantly in terms of the absorption of moisture. However, at this point we do not know where those differences may lie. Significance for the analysis of variance should be followed up with some form of multiple comparison like the **Least Significant Difference**.