

### A2. Statistics from a Geometric Viewpoint

**Analysis of Variance**

The technique of multiple comparison of treatment means via ANOVA can be viewed very elegantly, from a purely geometric perspective. Again, recall some basic facts from elementary vector analysis:

For any two column vectors \( \mathbf{v} = (v_1, v_2, \ldots, v_n)^T \) and \( \mathbf{w} = (w_1, w_2, \ldots, w_n)^T \) in \( \mathbb{R}^n \), the standard Euclidean *dot product* \( \mathbf{v} \cdot \mathbf{w} \) is defined as \( \mathbf{v}^T \mathbf{w} = \sum_{i=1}^{n} v_i w_i \), hence is a scalar. Technically, the dot product is a special case of a more general mathematical object known as an *inner product*, denoted by \( \langle \mathbf{v}, \mathbf{w} \rangle \), and these notations are often used interchangeably. The length, or *norm*, of a vector \( \mathbf{v} \) can therefore be characterized as \( \|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{\sum_{i=1}^{n} v_i^2} \), and the included angle \( \theta \) between two vectors \( \mathbf{v} \) and \( \mathbf{w} \) can be calculated via the formula

\[
\cos \theta = \frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\|\mathbf{v}\| \|\mathbf{w}\|}, \quad 0 \leq \theta \leq \pi.
\]

From this relation, it is easily seen that two vectors \( \mathbf{v} \) and \( \mathbf{w} \) are *orthogonal* (i.e., \( \theta = \pi/2 \)), written \( \mathbf{v} \perp \mathbf{w} \), if and only if their dot product is equal to zero, i.e., \( \langle \mathbf{v}, \mathbf{w} \rangle = 0 \).

Now suppose we have sample data from \( k \) treatment groups of sizes \( n_1, n_2, \ldots, n_k \), respectively, which we organize in vector form as follows:

<table>
<thead>
<tr>
<th>Treatment 1</th>
<th>Treatment 2</th>
<th>. . . . . .</th>
<th>Treatment k</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y_1 = \begin{pmatrix} y_{11} \ y_{12} \ y_{13} \ \vdots \ y_{1n_1} \end{pmatrix} )</td>
<td>( y_2 = \begin{pmatrix} y_{21} \ y_{22} \ y_{23} \ \vdots \ y_{2n_2} \end{pmatrix} )</td>
<td>. . . . . .</td>
<td>( y_k = \begin{pmatrix} y_{k1} \ y_{k2} \ y_{k3} \ \vdots \ y_{kn_k} \end{pmatrix} )</td>
</tr>
</tbody>
</table>

**Group Means:** \( \bar{y}_1 \), \( \bar{y}_2 \), . . . . . . , \( \bar{y}_k \)

**Group Variances:** \( s_1^2 \), \( s_2^2 \), . . . . . . , \( s_k^2 \)

**Grand Mean:**

\[
\bar{y} = \frac{n_1 \bar{y}_1 + n_2 \bar{y}_2 + \ldots + n_k \bar{y}_k}{n},
\]

where \( n = n_1 + n_2 + \ldots + n_k \) is the combined sample size.

**Pooled Variance:**

\[
s_{\text{within groups}}^2 = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2 + \ldots + (n_k - 1) s_k^2}{n - k}
\]
Now, for Treatment column $i = 1, 2, \ldots, k$ and row $j = 1, 2, \ldots, n_i$, it is clear from simple algebra that

$$y_{ij} - \bar{y} = (\bar{y}_i - \bar{y}) + (y_{ij} - \bar{y}_i).$$

Therefore, for each Treatment $i = 1, 2, \ldots, k$, we have the $n_i$-dimensional column vector identity

$$y_i - \bar{y} \mathbf{1} = (\bar{y}_i - \bar{y}) \mathbf{1} + (y_i - \bar{y}_i \mathbf{1}),$$

where the $n_i$-dimensional vector $\mathbf{1} = (1, 1, \ldots, 1)^T$. Hence, vertically stacking these $k$ columns produces a vector identity in $\mathbb{R}^n$.

<table>
<thead>
<tr>
<th>Treatment $i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1$</td>
</tr>
<tr>
<td>$2$</td>
</tr>
<tr>
<td>$3$</td>
</tr>
<tr>
<td>$\vdots$</td>
</tr>
<tr>
<td>$k$</td>
</tr>
</tbody>
</table>

or, more succinctly...

$$\mathbf{u} = \mathbf{v} + \mathbf{w}.$$

But the two vectors $\mathbf{v}$ and $\mathbf{w}$ are orthogonal, since they have a zero dot product:

$$\mathbf{v}^T \mathbf{w} = \sum_{i=1}^{k} (\bar{y}_i - \bar{y}) \mathbf{1}^T (y_i - \bar{y}_i \mathbf{1})$$

$$\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i) = 0,$$

because this is the sum of the deviations of the $y_{ij}$ values in Treatment $i$ from their group mean $\bar{y}_i$.

Therefore, the three vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ form a right triangle, as shown. So by the Pythagorean Theorem,
or, in statistical notation...

\[ \text{SS}_{\text{Total}} = \text{SS}_{\text{Trt}} + \text{SS}_{\text{Error}} \]

where

\[
\text{SS}_{\text{Total}} = \| \mathbf{u} \|^2 = \sum_{i=1}^{k} \left\| y_i - \bar{y} \right\|^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 = \sum_{\text{all } i, j} (y_{ij} - \bar{y})^2
\]

The sum of the squared deviations of each observation from the grand mean.

\[
\text{SS}_{\text{Trt}} = \| \mathbf{v} \|^2 = \sum_{i=1}^{k} \left\| \bar{y}_i - \bar{y} \right\|^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\bar{y}_i - \bar{y})^2 = \sum_{i=1}^{k} n_i (\bar{y}_i - \bar{y})^2
\]

The sum of the squared deviations of each group mean from the grand mean.

\[
\text{SS}_{\text{Error}} = \| \mathbf{w} \|^2 = \sum_{i=1}^{k} \left\| y_i - \bar{y}_i \right\|^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 = \sum_{i=1}^{k} (n_i - 1) s_i^2
\]

The sum of the squared deviations of each observation from its group mean.

The resulting ANOVA table for the null hypothesis \( H_0: \mu_1 = \mu_2 = \ldots = \mu_k \) is given by:

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>( F )-statistic</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>( k - 1 )</td>
<td>( \sum_{i=1}^{k} n_i (\bar{y}_i - \bar{y})^2 )</td>
<td>( s_{\text{between groups}}^2 )</td>
<td>( F_{k-1, n-k} )</td>
<td>( 0 \leq p \leq 1 )</td>
</tr>
<tr>
<td>Error</td>
<td>( n - k )</td>
<td>( \sum_{i=1}^{k} (n_i - 1) s_i^2 )</td>
<td>( s_{\text{within groups}}^2 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>( n - 1 )</td>
<td>( \sum_{\text{all } i, j} (y_{ij} - \bar{y})^2 )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
One final note about multiple treatment comparisons… We may also express the problem via the following equivalent formulation: For each Treatment column $i = 1, 2, \ldots, k$ and row $j = 1, 2, \ldots, n_i$, the $(i,j)^{th}$ response $y_{ij}$ differs from its true group mean $\mu_i$ by a random error amount $\varepsilon_{ij}$. At the same time however, the true group mean $\mu_i$ itself differs from the true grand mean $\mu$ by a random amount $\alpha_i$, appropriately called the $i^{th}$ treatment effect. That is,

$$y_{ij} = \mu_i + \varepsilon_{ij}$$

i.e.,

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

In words, this so-called model equation says that each individual response can be formulated as the sum of the grand mean plus its group treatment effect (the two of these together sum to its group mean), and an individual error term. The null hypothesis that “all of the group means are equal to each other” translates to the equivalent null hypothesis that “all of the group treatment effects are equal to zero.”

This expression of the problem as “response = model + error” is extremely useful, and will appear again, in the context of regression models.