COMPARING MORE THAN TWO POPULATION MEANS: AN ANALYSIS OF VARIANCE

To see how the principle behind the analysis of variance method works, let us consider the following simple experiment. The means ($\mu_1$ and $\mu_2$) of two populations are to be compared using independent random samples of size $n_1 = n_2 = 5$ from each of the populations. The sample observations and the sample means are shown below:

<table>
<thead>
<tr>
<th>Sample from Population 1</th>
<th>Sample from Population 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>$\bar{x}_1 = 2$</td>
<td>$\bar{x}_2 = 3$</td>
</tr>
</tbody>
</table>

Do you think these data provide sufficient evidence to indicate a difference between the population means $\mu_1$ and $\mu_2$?

Now look at two more samples of $n_1 = n_2 = 5$ measurements from the populations, as shown below. Do these data appear to provide evidence of a difference between $\mu_1$ and $\mu_2$?

<table>
<thead>
<tr>
<th>Sample from Population 1</th>
<th>Sample from Population 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>$\bar{x}_1 = 2$</td>
<td>$\bar{x}_2 = 3$</td>
</tr>
</tbody>
</table>

One way to determine whether a difference exists between the population means $\mu_1$ and $\mu_2$ is to examine the spread (or variation) between the sample means $\bar{x}_1$ and $\bar{x}_2$, and to compare it to a measure of variability within the samples. The greater the difference in the variations, the greater will be the evidence to indicate a difference between $\mu_1$ and $\mu_2$.

For the data in the first table, you can see that the difference between the sample means is small relative to the variability within the sample observations. Thus, we think you will agree that the difference between $\bar{x}_1$ and $\bar{x}_2$ is not large enough to indicate a difference between $\mu_1$ and $\mu_2$.

Notice that the difference between the sample means for the data in the second table is identical to the difference shown in the first table. However, since there is now no variability within the sample observations, the difference between the sample means is large compared to the variability within the sample observations. Thus, the data appear to give clear evidence of a difference between $\mu_1$ and $\mu_2$.

We apply the principle of this example to the general problem of comparing $k$ population means: If the variability among the $k$ sample means is large relative to the variability within the $k$ samples, then there is evidence to indicate that a difference exists among the $k$ population means.
ONE-WAY ANALYSIS OF VARIANCE

Suppose that we have \( r \) samples of respective sizes \( n_1, n_2, \ldots, n_r \). Let

\[
SS_b = \text{the between sample sum of squares}, \text{ which measures the variation between the sample means (also known as SST, the sum of squares for treatments, since it measures the variation due to the differences between treatments, or explained variation, since it measures variation that might be due to the inherent differences in the treatments)},
\]

\[
= \sum_{i=1}^{r} n_i (\bar{x}_{i*} - \bar{x}**)^2 = \sum_{i=1}^{r} n_i \bar{x}_{i*}^2 - N \bar{x}**^2
\]

where \( \bar{x}_{i*} \), the sample mean = \( \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \), \( i = 1, \ldots, r \)

\( \bar{x}** \), the overall mean = \( \frac{1}{N} \sum_{i=1}^{r} \sum_{j=1}^{n_i} x_{ij} \), where \( N = \sum_{i=1}^{r} n_i \)

\[
SS_w = \text{the within sample sum of squares, which measures the variation within the samples (also known as SSE, the sum of squares for errors, or unexplained variation, since it measures variation due to chance, for which no identifiable cause can be found)},
\]

\[
= \sum_{i=1}^{r} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i*})^2 = \sum_{i=1}^{r} \sum_{j=1}^{n_i} x_{ij}^2 - \sum_{i=1}^{r} n_i \bar{x}_{i*}^2
\]

Test statistic \( F_{r-1,N-r} = \frac{SS_b / (r - 1)}{SS_w / (N - r)} \) with numerator degree of freedom = \( r - 1 \), and denominator degree of freedom = \( N - r \).

Example

The chin-up scores of 12 children taking 3 types of diets are as follow:

<table>
<thead>
<tr>
<th></th>
<th>Chin-up scores</th>
<th>( \bar{x}_{1*} = 3.25 )</th>
<th>( \bar{x}** = 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal diet</td>
<td>3, 3, 3, 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>high protein diet</td>
<td>2, 4, 4, 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vegetarian diet</td>
<td>1, 2, 2, 3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Does the diet make a difference in the score obtained?

\[
SS_b = 4(3.25)^2 + 4(3.75)^2 + 4(2)^2 - 12(3)^2 = 114.5 - 108 = 6.5
\]

\[
SS_w = 3^2 + 3^2 + \ldots + 2^2 + 3^2 - 114.5 = 122 - 114.5 = 7.5
\]

Test statistic \( F_{2,9} = \frac{6.5 / 2}{7.5 / 9} = 3.9 \)

Since \( 3.9 < F_{0.05,2,9} = 4.26 \), we have insufficient evidence at 5% level of significance to conclude that the diet make a difference in the score obtained.
A Randomised Block Design is a design in which \( r \) treatments are compared within each of \( c \) blocks. Each block contains \( r \) matched experimental units and the \( r \) treatments are randomly assigned, one to each of the units within each block. Let

\[
SS_r = \text{the between row sum of squares which measures the variation between the row means (also known as SST, the sum of squares of treatments, since it measures the variation due to the differences between treatments)},
\]

\[
= c \sum_{i=1}^{r} (\bar{x}_i - \bar{x})^2 = \frac{1}{c} \sum_{j=1}^{c} \sum_{i=1}^{r} x_{ij}^2 - rc \bar{x}^2
\]

where \( \bar{x}_i \) , the mean of observations in row \( i \) = \( \frac{1}{c} \sum_{j=1}^{c} x_{ij} \), \( i = 1, \ldots, r \)

\[
\bar{x} \quad \text{, the overall mean = } \frac{1}{rc} \sum_{i=1}^{r} \sum_{j=1}^{c} x_{ij} = \frac{1}{r} \sum_{i=1}^{r} \bar{x}_i = \frac{1}{c} \sum_{j=1}^{c} \bar{x}_j
\]

\[
SS_c = \text{the between column sum of squares which measures the variation between the column means (also known as SSB, the sum of squares of blocks, since it measures the variation due to the differences between blocks)},
\]

\[
= r \sum_{j=1}^{c} (\bar{x}_j - \bar{x})^2 = \frac{1}{r} \sum_{i=1}^{r} \sum_{j=1}^{c} x_{ij}^2 - rc \bar{x}^2
\]

where \( \bar{x}_j \) , the mean of observations in column \( j \) = \( \frac{1}{r} \sum_{i=1}^{r} x_{ij} \), \( j = 1, \ldots, c \)

\[
SS_{\text{total}} = \text{the total sum of squares}
\]

\[
= \sum_{i=1}^{r} \sum_{j=1}^{c} (x_{ij} - \bar{x})^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} x_{ij}^2 - rc \bar{x}^2
\]

\[
SS_e = \text{the error sum of squares which measures the variation within the samples}
\]

\[
= \sum_{i=1}^{r} \sum_{j=1}^{c} (x_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2 = SS_{\text{total}} - SS_r - SS_c
\]

Test statistic for computing row effects \( F_{r-1,(r-1)(c-1)} = \frac{SS_r / (r-1)}{SS_e / (r-1)(c-1)} \) with numerator degree of freedom = \( r - 1 \), and denominator degree of freedom = \( (r-1)(c-1) \).

Test statistic for computing column effects \( F_{c-1,(r-1)(c-1)} = \frac{SS_c / (c-1)}{SS_e / (r-1)(c-1)} \) with numerator degree of freedom = \( c - 1 \), and denominator degree of freedom = \( (r-1)(c-1) \).
Example
The chin-up scores of four children when taking different types of diets are as follow:

<table>
<thead>
<tr>
<th>blocks</th>
<th>Alan</th>
<th>Bob</th>
<th>Carl</th>
<th>Dave</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal diet</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>high protein diet</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>vegetarian diet</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

\[ \bar{x}_{1*} = 3.25 \]
\[ \bar{x}_{2*} = 3.75 \]
\[ \bar{x}_{3*} = 3 \]

\[
SS_t = 4 \left(3.25^2 + 3.75^2 + 2^2\right) - 12 (3)^2 = 114.5 - 108 = 6.5
\]
\[
SS_c = 3 \left(2^2 + 3^2 + 3^2 + 4^2\right) - 12 (3)^2 = 114 - 108 = 6
\]
\[
SS_{total} = 3^2 + 3^2 + ... + 2^2 + 3^2 - 12 (3)^2 = 122 - 108 = 14
\]
\[
SS_e = SS_{total} - SS_t - SS_c = 14 - 6.5 - 6 = 1.5
\]

Test statistic for comparing effects of diet \( F_{2,6} = \frac{6.5 / 2}{1.5 / 6} = 13 \)

Test statistic for comparing effects of child \( F_{3,6} = \frac{6 / 3}{1.5 / 6} = 8 \)

Since 13 > \( F_{0.05,2,6} = 5.14 \), we have sufficient evidence at 5% level of significance to conclude that the diet does make a difference in the score obtained.

Since 8 > \( F_{0.05,3,6} = 4.76 \), we have sufficient evidence at 5% level of significance to conclude that the child does make a difference in the score obtained.

Note:
(1) Although it is possible to calculate a value of \( F \) for the columns (blocks), this variable is usually of a secondary interest and is used primarily to allow finer discrimination between rows (treatments).

(2) \( SS_e \) is smaller than in the earlier one-way analysis, because much of the formerly unexplained variation in the sample observations can now be attributed to differences in the children. We obtain a much higher value of \( F \) for rows (treatments) than before, allowing us to reject the null hypothesis (equal row/treatment population means) at a smaller level of significance than we could in the one-way analysis. Two-way analysis of variance is thus more efficient than one-way analysis.

(3) This method assumes that the row and column effects are additive. That is, each diet affected all children equally. We did not allow for the possibility that a particular diet may interact strongly with a particular child.
A Factorial Design is one conducted to investigate the effect of two factors on the mean value of a response variable. One factor has r levels and the other factor has c levels, and n (the number of replications) responses are measured for each of the r \times c factor level combinations. Let

$$SS_r = \text{the row sum of squares}$$

$$= cn \sum_{i=1}^{r} \left( \bar{x}_{\cdot \cdot i} - \bar{x}_{\cdot \cdot \cdot} \right)^2 = cn \sum_{i=1}^{r} \bar{x}_{\cdot \cdot i}^2 - rcn \bar{x}_{\cdot \cdot \cdot}^2$$

where \( \bar{x}_{\cdot \cdot i} \), the mean of observations in row i = \( \frac{1}{cn} \sum_{j=1}^{c} \sum_{k=1}^{n} x_{ijk} \), i = 1, ..., r

\( \bar{x}_{\cdot \cdot \cdot} \), the overall mean = \( \frac{1}{rcn} \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k=1}^{n} x_{ijk} \)

$$SS_c = \text{the column sum of squares}$$

$$= rn \sum_{j=1}^{c} \left( \bar{x}_{\cdot j \cdot} - \bar{x}_{\cdot \cdot \cdot} \right)^2 = rm \sum_{j=1}^{c} \bar{x}_{\cdot j \cdot}^2 - rcn \bar{x}_{\cdot \cdot \cdot}^2$$

where \( \bar{x}_{\cdot j \cdot} \), the mean of observations in column j = \( \frac{1}{rn} \sum_{i=1}^{r} \sum_{k=1}^{n} x_{ijk} \), j = 1, ..., c

$$SS_{\text{int}} = \text{the interaction sum of squares}$$

$$= n \sum_{i=1}^{r} \sum_{j=1}^{c} \left( \bar{x}_{ij \cdot} - \bar{x}_{\cdot \cdot i} - \bar{x}_{\cdot j \cdot} + \bar{x}_{\cdot \cdot \cdot} \right)^2 = n \sum_{i=1}^{r} \sum_{j=1}^{c} \left( \bar{x}_{ij \cdot} - \bar{x}_{\cdot \cdot \cdot} \right)^2 - SS_r - SS_c$$

$$= n \sum_{i=1}^{r} \sum_{j=1}^{c} \bar{x}_{ij \cdot}^2 - rcn \bar{x}_{\cdot \cdot \cdot}^2 - SS_r - SS_c$$

where \( \bar{x}_{ij \cdot} \), the mean of observations in row i, column j = \( \frac{1}{n} \sum_{k=1}^{n} x_{ijk} \)

$$SS_{\text{total}} = \text{the total sum of squares}$$

$$= \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k=1}^{n} (x_{ijk} - \bar{x}_{\cdot \cdot \cdot})^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k=1}^{n} x_{ijk}^2 - rcn \bar{x}_{\cdot \cdot \cdot}^2$$

$$SS_{\text{e}} = \text{the error sum of squares}$$

$$= \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k=1}^{n} (x_{ijk} - \bar{x}_{ij \cdot})^2 = SS_{\text{total}} - SS_r - SS_c - SS_{\text{int}}$$

Test statistic for detecting interaction \( F_{(r-1)(c-1),rc(n-1)} = \frac{SS_{\text{int}} / (r-1)(c-1)}{SS_{\text{e}} / rc(n-1)} \) with numerator degree of freedom = \( (r-1)(c-1) \), and denominator degree of freedom = \( rc(n-1) \).

Test statistic for computing row effects \( F_{r-1,rc(n-1)} = \frac{SS_r / (r-1)}{SS_{\text{e}} / rc(n-1)} \) with numerator degree of freedom = \( r-1 \), and denominator degree of freedom = \( rc(n-1) \).

Test statistic for computing column effects \( F_{c-1,rc(n-1)} = \frac{SS_c / (c-1)}{SS_{\text{e}} / rc(n-1)} \) with numerator degree of freedom = \( c-1 \), and denominator degree of freedom = \( rc(n-1) \).
Example
The following represents the chin-up scores of 6 girls and 6 boys taking 3 types of diets:

<table>
<thead>
<tr>
<th></th>
<th>girls</th>
<th>boys</th>
</tr>
</thead>
<tbody>
<tr>
<td>vegetarian</td>
<td>1 3</td>
<td>2 4</td>
</tr>
<tr>
<td>high protein</td>
<td>5 7</td>
<td>6 9</td>
</tr>
<tr>
<td>normal</td>
<td>7 4</td>
<td>9 6</td>
</tr>
</tbody>
</table>

Is there any evidence of an interaction effect? Do the data indicate that the diet and the sex of the child do indeed affect the score?

\[
\begin{align*}
\text{SS}_r &= 4(2.5^2 + 6.75^2 + 6.5^2) - 12(5.25)^2 = 45.5 \\
\text{SS}_c &= 6(5^2 + 5.5^2) - 12(5.25)^2 = 0.75 \\
\text{SS}_{\text{int}} &= 2(1.5^2 + 3.5^2 + 5.5^2 + 8^2 + 8^2 + 5^2) - 12(5.25)^2 - 45.5 - 0.75 = 18.5 \\
\text{SS}_{\text{total}} &= 1^2 + 2^2 + \ldots + 4^2 + 6^2 - 12(5.25)^2 = 72.25 \\
\text{SS}_e &= 72.25 - 45.5 - 0.75 - 18.5 = 7.5 \\
\end{align*}
\]

Test statistic for detecting interaction \( F_{2,6} = \frac{18.5 / 2}{7.5 / 6} = 7.4 \)
Since 7.4 > \( F_{0.05,2,6} = 5.14 \), we have sufficient evidence at 5% level of significance to conclude that there is interaction between the diet and the sex of the child.

Notice that the mean scores for girls are less than those for boys taking vegetarian or high protein diets. But the means are reversed for children taking a normal diet. Since the mean depends on the combination of the factor levels, we say that the two factors interact.

Note: The F tests for factor effects are usually relevant only when factor interaction is insignificant. If interaction is detected, we do not perform the F tests for factor effects. In fact, one of the most important objectives of a two-way analysis of variance with interaction is to detect factor interaction if it exists.
Example
Suppose now the scores are as shown below:

<table>
<thead>
<tr>
<th>Vegetarian Diet</th>
<th>Girls</th>
<th>Boys</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1,5</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>High Protein Diet</th>
<th>Girls</th>
<th>Boys</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>5,5</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Normal Diet</th>
<th>Girls</th>
<th>Boys</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>

Is there any evidence of an interaction effect? Do the data indicate that the diet and the sex of the child do indeed affect the score?

\[
\begin{align*}
SS_r &= 4(2.5^2 + 6.75^2 + 6.5^2) - 12(5.25)^2 = 45.5 \\
SS_c &= 6(4^2 + 6.5^2) - 12(5.25)^2 = 18.75 \\
SS_{int} &= 2(1.5^2 + 3.5^2 + 5.5^2 + 8^2 + 5^2 + 8^2) - 12(5.25)^2 - 45.5 - 18.75 = 0.5 \\
SS_{total} &= 1^2 + 2^2 + ... + 7^2 + 9^2 - 12(5.25)^2 = 72.25 \\
SS_e &= 72.25 - 45.5 - 18.75 - 0.5 = 7.5
\end{align*}
\]

Test statistic for detecting interaction \(F_{2,6} = \frac{0.5/2}{7.5/6} = 0.2\)

Since \(0.2 < F_{0.05,2,6} = 5.14\), we have insufficient evidence at 5% level of significance to conclude that there is interaction between the diet and the sex of the child.

We now focus on testing the effects of the diet and the sex of the child.

Test statistic for comparing effects of diet \(F_{2,6} = \frac{45.5/2}{7.5/6} = 18.2\)

Test statistic for comparing effects of sex of child \(F_{1,6} = \frac{18.75/1}{7.5/6} = 15\)

Since \(18.2 > F_{0.05,2,6} = 5.14\), we have sufficient evidence at 5% level of significance to conclude that the diet does make a difference in the score obtained.

Since \(15 > F_{0.05,1,6} = 5.99\), we have sufficient evidence at 5% level of significance to conclude that the sex of the child does make a difference in the score obtained.