Variance: \[ \frac{\sum (x - \bar{x})^2}{n-1} = \frac{SS_X}{df} \leq \text{mean squared}, MS \]

\(K = \# \text{ of groups}\)
\(j = \text{ group } \# \quad 1 \leq j \leq K\)
\(i = \text{ sample } \# \text{ within a group}\)

\(X_{ij} = i^{th} \text{ sample within group } j\quad X_{3,2} = 3^{rd} \text{ sample in the 2nd group}\)

\(n_j = \text{ sample size for group } j\)

\(\bar{X}_j = \text{ mean of group } j\)

\(\bar{X} = \text{ grand mean (mean of all samples)}\)

\(N = \text{ total sample size}\)

Data:

\[
\begin{array}{cccc}
X_{11} & X_{12} & \ldots & X_{1K} \\
X_{21} & X_{22} & \ldots & X_{2K} \\
\vdots & \vdots & \ddots & \vdots \\
X_{n1} & X_{n2} & \ldots & X_{nK}
\end{array}
\]

\(\bar{X}_{ij} = \bar{X}_{ij} \pm (a - b) \pm (c - b)\)

\[(X_{i\bar{j}} - \bar{X}) = (X_{i\bar{j}} - \bar{X}_j) + (\bar{X}_j - \bar{X})\]
\[
\sum_{i} \left( x_{ij} - \bar{x} \right)^2 = \sum_{j} \left[ \sum_{i} \left( x_{ij} - \bar{x}_j \right)^2 \right] + \sum_{i} \left( \bar{x}_j - \bar{x} \right)^2 \]

Sum of squares from grand mean

\[SS_{\text{total}} = SS_{\text{within}} + SS_{\text{between}}\]

\[
df_{\text{total}} = df_{\text{within}} + df_{\text{between}}
\]

\[
N - 1 = N - k + k - 1
\]

\[
MS_{\text{total}} = \frac{SS_{\text{total}}}{df_{\text{total}}} \quad MS_{\text{within}} = \frac{SS_{\text{within}}}{df_{\text{within}}} \quad MS_{\text{between}} = \frac{SS_{\text{between}}}{df_{\text{between}}}
\]

\[
F = \frac{MS_{\text{between}}}{MS_{\text{within}}}
\]
Example 2: Preferred temperature for weather sensitivity

At the beginning of the quarter I surveyed you for your preferred outdoor temperature. I also asked you how much weather affected your mood with the options of Not at all, Just a little, A fair amount and Very much. Let’s see if there is a significant difference between the preferred temperatures across these 4 options. We’ll use α = 0.05 again. Here’s a table of statistics:

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>Just a little</th>
<th>A fair amount</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>12</td>
<td>35</td>
<td>63</td>
<td>40</td>
</tr>
<tr>
<td>mean</td>
<td>69.13</td>
<td>69.53</td>
<td>70.98</td>
<td>73.62</td>
</tr>
<tr>
<td>SS</td>
<td>701.5468</td>
<td>2770.8755</td>
<td>2612.9852</td>
<td>1490.838</td>
</tr>
<tr>
<td>s</td>
<td>7.986</td>
<td>9.0275</td>
<td>6.4919</td>
<td>6.1828</td>
</tr>
<tr>
<td>sem</td>
<td>2.3054</td>
<td>1.5259</td>
<td>0.8179</td>
<td>0.9776</td>
</tr>
</tbody>
</table>

Totals:

<table>
<thead>
<tr>
<th></th>
<th>150</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td></td>
</tr>
<tr>
<td>grand mean</td>
<td>71.1993</td>
</tr>
<tr>
<td>SS\text{total}</td>
<td>7961.6499</td>
</tr>
</tbody>
</table>

\[
\frac{\sqrt{701.5468}}{\sqrt{n}} = \frac{7.986}{\sqrt{12}} = 2.3054
\]

Numerator of the F statistic:

\[
\text{SS}_\text{between} = \frac{\text{SS}_\text{between}}{\text{df}_\text{between}} = \frac{150 - 1 = 4 - 1 = 3}{3}
\]

\[
\text{SS}_\text{between} = \sum n_j \left( \bar{x}_j - \bar{x} \right)^2
\]

\[
= Q \left( 69.13 - 71.1993 \right)^2 + 35 \left( 69.53 - 71.1993 \right)^2 + \ldots
\]

\[
= 385.4044
\]

\[
\text{MS}_\text{between} = \frac{385.4044}{3} = 128.7784
\]

Denominator:

\[
\text{SS}_\text{within} = \frac{\text{SS}_\text{within}}{\text{df}_\text{within}}
\]

\[
\text{SS}_\text{within} = 701.5468 + 2770.8755 + \ldots = 7876.2455
\]

\[
\text{df}_\text{within} = N - k = 150 - 4 = 146
\]
\[
\text{MS within} = \frac{2576.2455}{146} = 17.8921
\]

\[
F = \frac{\text{MS between}}{\text{MS within}} = \frac{128.7784}{17.8921} = 7.24817.
\]