Using proportions to calculate the mean

In this calculation the notation has a way of making the everything look more complicated than it is, so let’s start with an example.

Suppose the list is 7, 2, 2, 7, 8, 3, 2, 7, 8, 8. Then there are 10 entries, among which there are 4 distinct values: 2, 3, 7, and 8. Organize the list in a distribution table:

<table>
<thead>
<tr>
<th>value</th>
<th>number of entries</th>
<th>proportion of entries</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3</td>
<td>( \frac{3}{10} )</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>( \frac{1}{10} )</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>( \frac{3}{10} )</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>( \frac{3}{10} )</td>
</tr>
<tr>
<td></td>
<td>total = 10</td>
<td>total = 1</td>
</tr>
</tbody>
</table>

The mean is the sum of all the entries, divided by 10. The table gives us a quick way of calculating the sum. The mean equals

\[
\frac{2 \times 3 + 3 \times 1 + 7 \times 3 + 8 \times 3}{10} = 2 \times \frac{3}{10} + 3 \times \frac{1}{10} + 7 \times \frac{3}{10} + 8 \times \frac{3}{10}
\]

That works out to 5.4. But the main point here is that the mean can be calculated using only the distinct values and their proportions.

**The general calculation.** Now the algebra should be clear. Suppose that among the \( n \) entries in your list there are \( k \) distinct values. Here \( k \) can be any number between 1 and \( n \).

Call the distinct values \( v_1, v_2, \ldots, v_k \). Suppose the value \( v_1 \) appears \( n_1 \) times, \( v_2 \) appears \( n_2 \) times, and so on. Then for each \( i \) in the range 1 through \( k \), then the proportion of entries that have the value \( v_i \) is

\[
p_i = \frac{n_i}{n}
\]

The mean is

\[
\frac{1}{n} \sum_{i=1}^{k} v_i n_i = \sum_{i=1}^{k} \frac{v_i n_i}{n} = \sum_{i=1}^{k} v_i p_i
\]

The number of terms in the sum is the number of distinct values \( (k) \), not the total number of entries \( (n) \).