Correlation

Robert S Michael

The correlation coefficient can be calculated in several different ways. The first method, illustrated below, is the traditional deviation method used routinely in the days before computers, and used today to illustrate the concept of correlation. First, we begin with a set of two scores for each case:

<table>
<thead>
<tr>
<th>case</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11</td>
<td>75</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>72</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>75</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>63</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
<td>67</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td>67</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>52</td>
</tr>
<tr>
<td>8</td>
<td>15</td>
<td>68</td>
</tr>
<tr>
<td>9</td>
<td>18</td>
<td>60</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>68</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>75</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>72</td>
</tr>
<tr>
<td>13</td>
<td>11</td>
<td>67</td>
</tr>
<tr>
<td>14</td>
<td>11</td>
<td>54</td>
</tr>
<tr>
<td>15</td>
<td>11</td>
<td>41</td>
</tr>
<tr>
<td>16</td>
<td>21</td>
<td>90</td>
</tr>
</tbody>
</table>

To gain an impression of the degree of the relationship that may exist between two sets of scores, sort the scores on either variable and observe the sequence in the other variable.
Deviation Method

The steps for calculating the correlation coefficient by the deviation method are:

1. Find the sum of the X values ($\Sigma X$) and the sum of the Y values ($\Sigma Y$). Find the mean of the X values ($\overline{X}$) and the mean of the Y values ($\overline{Y}$).

<table>
<thead>
<tr>
<th>case</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11</td>
<td>75</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>72</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>75</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>63</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
<td>67</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>67</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>52</td>
</tr>
<tr>
<td>8</td>
<td>15</td>
<td>68</td>
</tr>
<tr>
<td>9</td>
<td>18</td>
<td>60</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>68</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>75</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>72</td>
</tr>
<tr>
<td>13</td>
<td>11</td>
<td>67</td>
</tr>
<tr>
<td>14</td>
<td>11</td>
<td>54</td>
</tr>
<tr>
<td>15</td>
<td>11</td>
<td>41</td>
</tr>
<tr>
<td>16</td>
<td>21</td>
<td>90</td>
</tr>
</tbody>
</table>

Sum = 205
Mean = 12.81

Table 0.2: Sums and Means for X and Y scores

Sum = 1066
Mean = 66.63
2 Find the deviations from the respective means for each X value \((X - \bar{X})\) and for each Y value \((Y - \bar{Y})\).

<table>
<thead>
<tr>
<th>case</th>
<th>X</th>
<th>X - (\bar{X})</th>
<th>Y</th>
<th>Y - (\bar{Y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11</td>
<td>-1.81</td>
<td>75</td>
<td>8.38</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>2.19</td>
<td>72</td>
<td>5.38</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>-0.81</td>
<td>75</td>
<td>8.38</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>1.19</td>
<td>63</td>
<td>-3.63</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
<td>-1.81</td>
<td>67</td>
<td>0.38</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>-3.81</td>
<td>67</td>
<td>0.38</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>-1.81</td>
<td>52</td>
<td>-14.63</td>
</tr>
<tr>
<td>8</td>
<td>15</td>
<td>2.19</td>
<td>68</td>
<td>1.38</td>
</tr>
<tr>
<td>9</td>
<td>18</td>
<td>5.19</td>
<td>60</td>
<td>-6.63</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>-0.81</td>
<td>68</td>
<td>1.38</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>-1.81</td>
<td>75</td>
<td>8.38</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>-0.81</td>
<td>72</td>
<td>5.38</td>
</tr>
<tr>
<td>13</td>
<td>11</td>
<td>-1.81</td>
<td>67</td>
<td>0.38</td>
</tr>
<tr>
<td>14</td>
<td>11</td>
<td>-1.81</td>
<td>54</td>
<td>-12.63</td>
</tr>
<tr>
<td>15</td>
<td>11</td>
<td>-1.81</td>
<td>41</td>
<td>-25.63</td>
</tr>
<tr>
<td>16</td>
<td>21</td>
<td>8.19</td>
<td>90</td>
<td>23.38</td>
</tr>
</tbody>
</table>

| Sum =  | 205 | 0.00 | Sum =  | 1066 | 0.00 |
| Mean = | 12.81 | Mean = | 66.63 |
| Sum \(X^2\) = | 2,771 | Sum \(Y^2\) = | 72,928 |
Find the sum of the squared deviations, respectively, from the mean for the X values ($\Sigma(X - \bar{X})^2$) and the Y values ($\Sigma(Y - \bar{Y})^2$).

Find the variance and the standard deviation for each set of scores. Recall variance is the mean of the sum of squared deviations ($\text{var} = \frac{\Sigma SS}{N}$) and the standard deviation is simply the square root of the variance ($\text{SD} = \sqrt{\text{var}}$).

$(Y - \bar{Y})^2$

Table 0.4: Sum of squared deviations from mean of X and mean of Y scores

<table>
<thead>
<tr>
<th>case</th>
<th>X</th>
<th>X - X</th>
<th>$(X - \bar{X})^2$</th>
<th>Y</th>
<th>Y - Y</th>
<th>$(Y - \bar{Y})^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11</td>
<td>-1.81</td>
<td>3.29</td>
<td>75</td>
<td>8.38</td>
<td>70.14</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>2.19</td>
<td>4.79</td>
<td>72</td>
<td>5.38</td>
<td>28.89</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>-0.81</td>
<td>0.66</td>
<td>75</td>
<td>8.38</td>
<td>70.14</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>1.19</td>
<td>1.41</td>
<td>63</td>
<td>-3.63</td>
<td>13.14</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
<td>-1.81</td>
<td>3.29</td>
<td>67</td>
<td>0.38</td>
<td>0.14</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>-3.81</td>
<td>14.54</td>
<td>67</td>
<td>0.38</td>
<td>0.14</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>-1.81</td>
<td>3.29</td>
<td>52</td>
<td>-14.63</td>
<td>213.89</td>
</tr>
<tr>
<td>8</td>
<td>15</td>
<td>2.19</td>
<td>4.79</td>
<td>68</td>
<td>1.38</td>
<td>1.89</td>
</tr>
<tr>
<td>9</td>
<td>18</td>
<td>5.19</td>
<td>26.91</td>
<td>60</td>
<td>-6.63</td>
<td>43.89</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>-0.81</td>
<td>0.66</td>
<td>68</td>
<td>1.38</td>
<td>1.89</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>-1.81</td>
<td>3.29</td>
<td>75</td>
<td>8.38</td>
<td>70.14</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>-0.81</td>
<td>0.66</td>
<td>72</td>
<td>5.38</td>
<td>28.89</td>
</tr>
<tr>
<td>13</td>
<td>11</td>
<td>-1.81</td>
<td>3.29</td>
<td>67</td>
<td>0.38</td>
<td>0.14</td>
</tr>
<tr>
<td>14</td>
<td>11</td>
<td>-1.81</td>
<td>3.29</td>
<td>54</td>
<td>-12.63</td>
<td>159.39</td>
</tr>
<tr>
<td>15</td>
<td>11</td>
<td>-1.81</td>
<td>3.29</td>
<td>41</td>
<td>-25.63</td>
<td>656.64</td>
</tr>
<tr>
<td>16</td>
<td>21</td>
<td>8.19</td>
<td>67.04</td>
<td>90</td>
<td>23.38</td>
<td>546.39</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>205</th>
<th>0.00</th>
<th>144.44$^a$</th>
<th>1066</th>
<th>0.00</th>
<th>1905.75$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum</td>
<td>Mean</td>
<td></td>
<td></td>
<td>Sum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X$^2$</td>
<td>2,771</td>
<td>3.00$^f$</td>
<td>72,928</td>
<td>10.91$^f$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Sum of Squared Deviations for X
b. Sum of squares for Y
c. Variance for X (i.e., mean of the sum of the squared deviations)
d. Variance for Y
e. Standard Deviation of X (the square root of the variance)
f. Standard Deviation of Y
5 Find the cross products of the deviations from the respective means. That is, multiply each \( X - \bar{X} \) deviation by its corresponding \( Y - \bar{Y} \) deviation. Then sum the productions, \( \sum (X - \bar{X})(Y - \bar{Y}) \).

\[
(X - \bar{X})(Y - \bar{Y})
\]

Table 0.5: Sums and Means for X and Y scores

<table>
<thead>
<tr>
<th>case</th>
<th>X</th>
<th>( X - \bar{X} )</th>
<th>( (X - \bar{X})^2 )</th>
<th>Y</th>
<th>( Y - \bar{Y} )</th>
<th>( (Y - \bar{Y})^2 )</th>
<th>( (X - \bar{X})(Y - \bar{Y}) )</th>
<th>( X \cdot Y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11</td>
<td>-1.81</td>
<td>3.29</td>
<td>75</td>
<td>8.38</td>
<td>70.14</td>
<td>-15.18</td>
<td>825</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>2.19</td>
<td>4.79</td>
<td>72</td>
<td>5.38</td>
<td>28.89</td>
<td>11.76</td>
<td>1080</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>-0.81</td>
<td>0.66</td>
<td>75</td>
<td>8.38</td>
<td>70.14</td>
<td>-6.80</td>
<td>900</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>1.19</td>
<td>1.41</td>
<td>63</td>
<td>-3.63</td>
<td>13.14</td>
<td>-4.30</td>
<td>882</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
<td>-1.81</td>
<td>3.29</td>
<td>67</td>
<td>0.14</td>
<td>0.38</td>
<td>-0.68</td>
<td>737</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>-3.81</td>
<td>14.54</td>
<td>67</td>
<td>0.38</td>
<td>0.14</td>
<td>-1.43</td>
<td>603</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>-1.81</td>
<td>3.29</td>
<td>52</td>
<td>-14.63</td>
<td>213.89</td>
<td>26.51</td>
<td>572</td>
</tr>
<tr>
<td>8</td>
<td>15</td>
<td>2.19</td>
<td>4.79</td>
<td>68</td>
<td>1.38</td>
<td>1.89</td>
<td>1.89</td>
<td>1020</td>
</tr>
<tr>
<td>9</td>
<td>18</td>
<td>5.19</td>
<td>26.91</td>
<td>60</td>
<td>-6.63</td>
<td>43.89</td>
<td>-34.37</td>
<td>1080</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>-0.81</td>
<td>0.66</td>
<td>68</td>
<td>1.38</td>
<td>1.89</td>
<td>-1.12</td>
<td>816</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>-1.81</td>
<td>3.29</td>
<td>75</td>
<td>8.38</td>
<td>70.14</td>
<td>-15.18</td>
<td>825</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>-0.81</td>
<td>0.66</td>
<td>72</td>
<td>5.38</td>
<td>28.89</td>
<td>-4.37</td>
<td>864</td>
</tr>
<tr>
<td>13</td>
<td>11</td>
<td>-1.81</td>
<td>3.29</td>
<td>67</td>
<td>0.38</td>
<td>0.14</td>
<td>-0.68</td>
<td>737</td>
</tr>
<tr>
<td>14</td>
<td>11</td>
<td>-1.81</td>
<td>3.29</td>
<td>54</td>
<td>-12.63</td>
<td>159.39</td>
<td>22.88</td>
<td>594</td>
</tr>
<tr>
<td>15</td>
<td>11</td>
<td>-1.81</td>
<td>3.29</td>
<td>41</td>
<td>-25.63</td>
<td>656.64</td>
<td>46.45</td>
<td>451</td>
</tr>
<tr>
<td>16</td>
<td>21</td>
<td>8.19</td>
<td>67.04</td>
<td>90</td>
<td>23.38</td>
<td>546.39</td>
<td>191.38</td>
<td>1890</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sum</th>
<th>205</th>
<th>0.00</th>
<th>144.44</th>
<th>1066</th>
<th>0.00</th>
<th>1905.75</th>
<th>217.88</th>
<th>13,876</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>12.81</td>
<td></td>
<td>66.63</td>
<td>119.11</td>
<td>66.63</td>
<td>119.11</td>
<td>119.11</td>
<td>119.11</td>
</tr>
<tr>
<td>Sum X²</td>
<td>2,771</td>
<td>St Dev X=</td>
<td>3.00</td>
<td>72,928</td>
<td>StDev Y=</td>
<td>10.91</td>
<td>10.91</td>
<td>10.91</td>
</tr>
</tbody>
</table>

6 Put it all together:

\[
r = \frac{(X - \bar{X})(Y - \bar{Y})}{\sqrt{\text{VAR}_x \cdot \text{VAR}_y}} = \frac{217.88}{\sqrt{9.03 \cdot 119.11}} = \frac{217.88}{\sqrt{1075.56}} = \frac{217.88}{32.79} = \frac{217.88}{524.73} = 0.415
\]

If you have a calculator, fewer key strokes are needed for the following formula which can be shown to be equivalent to the preceding deviation formula:

\[
r = \frac{N \cdot \Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\left[ N \cdot \Sigma X^2 - (\Sigma X)^2 \right] \left[ N \cdot \Sigma Y^2 - (\Sigma Y)^2 \right]}}
\]
Z-Score Method

Conceptually, the correlation coefficient may be easier to understand by inspecting the following formula:

\[
 r = \frac{\sum z_x \cdot z_y}{N}
\]

Table 0.6: Sums and Means for X and Y scores; Sum of \(X^2\) and \(Y^2\)

<table>
<thead>
<tr>
<th>case</th>
<th>X</th>
<th>(Z_x)</th>
<th>Y</th>
<th>(Z_y)</th>
<th>(Z_x Z_y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11</td>
<td>-0.60</td>
<td>75</td>
<td>0.77</td>
<td>-0.462924</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>0.73</td>
<td>72</td>
<td>0.49</td>
<td>0.35857</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>-0.27</td>
<td>75</td>
<td>0.77</td>
<td>-0.207518</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>0.40</td>
<td>63</td>
<td>-0.33</td>
<td>-0.131277</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
<td>-0.60</td>
<td>67</td>
<td>0.03</td>
<td>-0.020728</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>-1.27</td>
<td>67</td>
<td>0.03</td>
<td>-0.043600</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>-0.60</td>
<td>52</td>
<td>-1.34</td>
<td>0.80839</td>
</tr>
<tr>
<td>8</td>
<td>15</td>
<td>0.73</td>
<td>68</td>
<td>0.13</td>
<td>0.091727</td>
</tr>
<tr>
<td>9</td>
<td>18</td>
<td>1.73</td>
<td>60</td>
<td>-0.61</td>
<td>-1.048072</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>-0.27</td>
<td>68</td>
<td>0.13</td>
<td>-0.034070</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>-0.60</td>
<td>75</td>
<td>0.77</td>
<td>-0.462924</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>-0.27</td>
<td>72</td>
<td>0.49</td>
<td>-0.133183</td>
</tr>
<tr>
<td>13</td>
<td>11</td>
<td>-0.60</td>
<td>67</td>
<td>0.03</td>
<td>-0.020728</td>
</tr>
<tr>
<td>14</td>
<td>11</td>
<td>-0.60</td>
<td>54</td>
<td>-1.16</td>
<td>0.697841</td>
</tr>
<tr>
<td>15</td>
<td>11</td>
<td>-0.60</td>
<td>41</td>
<td>-2.35</td>
<td>1.416410</td>
</tr>
<tr>
<td>16</td>
<td>21</td>
<td>2.73</td>
<td>90</td>
<td>2014</td>
<td>5.836466</td>
</tr>
</tbody>
</table>

Sum = 205 Sum = 1066 6.64438
Mean = 12.81 Mean = 66.63
StDev = 3.00 StDev = 10.91

\[
 r = \frac{\sum (z_x \cdot z_y)}{N} = \frac{6.64438}{16} = 0.4152
\]