Minimum and Maximum; Standard Deviation and Variance

Introduction to Excel

Limitations of the average

• Enter these two blocks of data:

<table>
<thead>
<tr>
<th>1</th>
<th>5</th>
<th>9</th>
<th>13</th>
<th>17</th>
<th>21</th>
<th>25</th>
<th>27</th>
<th>23</th>
<th>19</th>
<th>15</th>
<th>11</th>
<th>7</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>12</td>
<td>12</td>
<td>13</td>
<td>13</td>
<td>13.5</td>
<td>14</td>
<td>14</td>
<td>14.5</td>
<td>15</td>
<td>15</td>
<td>16</td>
<td>16</td>
<td>17</td>
</tr>
</tbody>
</table>

• How do their averages compare?

• Are they the same populations?
Comparing Datasets

- Average isn’t good enough
  - Two sets can have the same average, but be quite different
  - Two sets could be very similar, but have different averages

- The range and distribution also matter
  - Are all the values close together? Or are they spread far apart?

- **Minimum** and **maximum** are measure's of a dataset's range

- **Variance** and **standard deviation** are measurements of a dataset’s distribution

Minimum and Maximum

- `min()` function returns the smallest value in a range of values
  - Which is smaller: -100 or -3 ?

- `max()` function returns the largest value in a range of values

- `max() – min()` gives the range of a set of values
Standard deviation

- Standard deviation is a distance, from the average value, within which you find about 2/3 of the sample data points
  - These data sets each have 14 points, so about 2/3 of the data points is 10 points
  - Each has an average of 14
  - How far do you have to go from 14 (plus and minus) to include 10 data points, in each case?

```
 1  3  5  7  9  11  13  15  17  19  21  23  25  27
```

```
11 12 12 13 13 13.5 14 14 14.5 15 15 16 16 17
```

Calculating Standard Deviation and Variance

- Mathematically the variance is the square of the standard deviation
  - or,

\[
\text{std. dev.} = \sqrt{\text{variance}}
\]

- Excel has functions for both of these values
  - **stdev** (range) calculates the standard deviation of a range of cells
  - **var** (range) calculates the variance of a range of cells
### Practice: add standard deviation calculations to this spreadsheet

<table>
<thead>
<tr>
<th>Mon</th>
<th>Tues</th>
<th>Wed</th>
<th>Thurs</th>
<th>Fri</th>
<th>Sat</th>
<th>Sun</th>
<th>Sum</th>
<th>Count</th>
<th>Average</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>29</td>
<td>16</td>
<td>13</td>
<td>9</td>
<td>12</td>
<td>19</td>
<td>113</td>
<td>7</td>
<td>16.14</td>
<td>16.14</td>
</tr>
<tr>
<td>13</td>
<td>26</td>
<td>22</td>
<td>13</td>
<td>22</td>
<td>26</td>
<td>27</td>
<td>149</td>
<td>7</td>
<td>21.29</td>
<td>21.29</td>
</tr>
<tr>
<td>11</td>
<td>16</td>
<td>10</td>
<td>13</td>
<td>24</td>
<td>29</td>
<td>14</td>
<td>117</td>
<td>7</td>
<td>16.71</td>
<td>16.71</td>
</tr>
<tr>
<td>11</td>
<td>15</td>
<td>27</td>
<td>28</td>
<td>9</td>
<td>24</td>
<td>25</td>
<td>139</td>
<td>7</td>
<td>19.86</td>
<td>19.86</td>
</tr>
<tr>
<td>13</td>
<td>16</td>
<td>15</td>
<td>25</td>
<td>10</td>
<td>14</td>
<td>14</td>
<td>107</td>
<td>7</td>
<td>15.29</td>
<td>15.29</td>
</tr>
<tr>
<td>12</td>
<td>29</td>
<td>26</td>
<td>21</td>
<td>11</td>
<td>26</td>
<td>22</td>
<td>147</td>
<td>7</td>
<td>21.00</td>
<td>21.00</td>
</tr>
<tr>
<td>8</td>
<td>12</td>
<td>23</td>
<td>10</td>
<td>8</td>
<td>12</td>
<td>12</td>
<td>85</td>
<td>7</td>
<td>12.14</td>
<td>12.14</td>
</tr>
<tr>
<td>12</td>
<td>10</td>
<td>25</td>
<td>29</td>
<td>12</td>
<td>10</td>
<td>6</td>
<td>104</td>
<td>7</td>
<td>14.86</td>
<td>14.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>961</td>
<td>56</td>
<td>17.16</td>
<td></td>
</tr>
</tbody>
</table>

- This spreadsheet is just a copy of an earlier one.

### Kinds of Standard Deviation

- The calculation of variance changes, depending on whether your data set is all possible data (the "population") or just a representative subset (a "sample").
- Two standard-deviation functions (and two variance functions), based on population and on sample.
  - `stdev.p()` and `var.p()`
    - Assume the data points are all possible data
  - `stdev.s()` and `var.s()`
    - Assume the data points represent a sample of the population
    - `stdev()` is a synonym for `stdev.s()`
more on variance and standard deviation later...

Summary – Basic Statistic Functions

• These Excel functions all are used in the same way, to calculate useful statistics about a single set of data:
  - average()
  - count()
  - max()
  - min()
  - stdev()
    » stdev.s(), stdev.p()
  - sum()
  - var()
    » var.s(), var.p()

Remember, the "Insert Function" link (looks like "fx", above the cells) gives a list of all available functions, and help in using them.