A **STATISTIC** is a measure obtained from a sample. The notation for statistics uses Roman letters like \( x, s, p, r \). The total number of data values for a sample is labeled lowercase \( n \).

A **PARAMETER** is a measure obtained from a population. The notation for parameters uses Greek letters like \( \mu \) (mu), \( \sigma \) (sigma), \( \rho \) (rho). The total number of data values for a population is labeled uppercase \( N \).

For both statistics and population formulas, the individual data values are represented by \( X \)'s. The capital Greek letter sigma, \( \Sigma \), means to add, or find the sum. \( \sum X \) means to add all of the data values for a sample or population.

The **MEAN** (aka the arithmetic average) of a data set is found by adding all the data values and then dividing by the total number of data values.

For a sample, the mean is labeled \( \bar{X} \), read as "\( x \) bar". For a population, the mean is labeled \( \mu \).

\[
\bar{X} = \frac{\sum X}{N} \quad \text{THIS MEANS ADD UP ALL THE DATA VALUES, THEN DIVIDE BY THE NUMBER OF DATA VALUES.}
\]

\[
\mu = \frac{\sum X}{N} \quad \text{THEIR DIVIDE BY THE NUMBER OF DATA VALUES.}
\]

**Assume data is from a sample unless otherwise specified.**

**Rounding rules:** Do not round until the final answer is calculated. For mean, variance and standard deviation, round to one more decimal place than the original data.

Find the mean for the data set of the number of students enrolled in P. Conrad’s Beginning Algebra Part 2 classes. This data set is a convenience sample representing enrollment for all such Metro classes.

<table>
<thead>
<tr>
<th>21</th>
<th>17</th>
<th>18</th>
<th>21</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>2</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>23</td>
<td>21</td>
<td>20</td>
<td>14</td>
</tr>
</tbody>
</table>

Find the mean for the data set of the number of students enrolled in P. Conrad’s Beginning Algebra Part 2 classes. This data set is a convenience sample representing enrollment for all such Metro classes.

\[
\bar{X} = \frac{\sum X}{n} = \frac{223}{12} = 18.58\overline{3}
\]

Round to 18.6
To find an approximation for the mean of grouped data use:

\[
\bar{X} = \frac{\sum f \cdot X_m}{n}
\]

where \( f \) is the frequency and \( X_m \) is the midpoint for each class.

From Exercises 2 – 3, #18, LPGA Scores

<table>
<thead>
<tr>
<th>Score</th>
<th>( X_m )</th>
<th>( f )</th>
<th>( f \cdot X_m )</th>
</tr>
</thead>
<tbody>
<tr>
<td>202 – 204</td>
<td>203</td>
<td>2</td>
<td>406</td>
</tr>
<tr>
<td>205 – 207</td>
<td>206</td>
<td>7</td>
<td>144</td>
</tr>
<tr>
<td>208 – 210</td>
<td>209</td>
<td>16</td>
<td>334</td>
</tr>
<tr>
<td>211 – 213</td>
<td>212</td>
<td>26</td>
<td>551</td>
</tr>
<tr>
<td>214 – 216</td>
<td>215</td>
<td>18</td>
<td>387</td>
</tr>
<tr>
<td>217 – 219</td>
<td>218</td>
<td>4</td>
<td>872</td>
</tr>
</tbody>
</table>

\[
\text{TOTAL } \sum f = 73
\]

\[
\text{TOTAL } \sum f \cdot X_m = 15446
\]

\[
\bar{X} = \frac{15446}{73} = 211.586
\]

Rounded to 211.6

Some properties of the mean of a data set:

The mean is not necessarily one of the data values.

Since every data value is used to compute the mean, it is affected by extremely high and low values (outliers).

The mean varies less than the median or mode when they are computed for different samples from the same population.

The mean cannot be calculated for open-ended distributions.
The **median (MD)** of a data set is found by putting the data in numerical order and finding the midpoint of the list. The ordered data is called a data array. If there are an even number of data values, add the middle two values and divide by 2 to find the median.

For the previous data set: array: 2, 14, 17, 18, 19, 20, 21, 21, 21, 22, 23, 25

\[
\text{Median} = \frac{21 + 21}{2} = 20.5
\]

The middle is between 20 and 21.

Some properties of the median of a data set:

- If there is an even number of data values, the median is not necessarily one of the data values.
- The median is used to determine if a data value is in the upper or lower half of a distribution.
- The median is affected less than the mean by very high or low data values.

The **mode** of a data set is the value that occurs most often. A data set with one mode is called unimodal. If two values occur with the same greatest frequency, both are considered to be modes and the data set is called bimodal. If no data value occurs more than once, the set has no mode. For above, the mode is 21.

For a grouped frequency distribution, the class with the highest frequency is called the **modal class**. For the LPGA scores, the modal class is 211—213.

Some properties of the mode:

- The mode represents the most typical data value.
- The mode can be found for qualitative (non-numeric) variables measured at the nominal level.
- If a mode exists, it is one or more of the data values.

The **midrange (MR)** of a data set is found by adding the highest and lowest data values, then dividing by 2. This is a rough estimate of the mean which can be computed very quickly. It is affected by extremely high and low data values.

The midrange for the above data set is \[
\frac{25 + 2}{2} = \frac{27}{2} = 13.5
\]
For the Dist. of Stats Quarter Grades,
\[
\bar{x} = 82.5 \\
\text{Median} = 83 \\
\text{Mode} = 78? \\
\text{MR} = 77
\]

From Grouped Data, Mean = 82.4

For Enrollment
\[
\bar{x} = 18.6 \\
\text{Median} = 20.5 \\
\text{Mode} = 21 \\
\text{MR} = 13.5
Section 3-3 Measures of Variation

Measures of variation describe how variable, or "spread out" a data set is.

RANGE = highest value – lowest value

This is the simplest measure of variation. It is greatly affected by outliers.

VARIANCE = the average of the squares of the distance each data value is from the mean.

Population:
\[ \sigma^2 = \frac{\sum (X - \mu)^2}{N} \]
Sample:
\[ s^2 = \frac{\sum (X - \overline{X})^2}{n-1} \]

STANDARD DEVIATION = the square root of the variance.

Population:
\[ \sigma = \sqrt{\frac{\sum (X - \mu)^2}{N}} \]
Sample:
\[ s = \sqrt{\frac{\sum (X - \overline{X})^2}{n-1}} \]

Variance and standard deviation are computed using every data value in the set, so they are less affected by outliers. These measures represent how closely the data values are grouped about the mean.

Rounding rule for variance and standard deviation: the final answer should be rounded to one more decimal place than the original data.

Variance for grouped data (sample):
\[ s^2 = \frac{\sum f \cdot X_m^2 - \left( \frac{\sum f \cdot X_m}{n} \right)^2}{n-1} \]

Use the COEFFICIENT OF VARIATION to compare the variation of data sets with different units:

population \( CVar = \frac{\sigma}{\mu} \times 100\% \)

sample \( CVar = \frac{s}{X} \times 100\% \)

Use the range rule of thumb, \( s \approx \frac{\text{range}}{4} \), for a rough estimate of the standard deviation.
Example: Find the range, variance, and standard deviation for these sample scores from two tests.

**Test 1**  \( \bar{X} = 82.0 \text{ points} \)

<table>
<thead>
<tr>
<th>( X )</th>
<th>( X - \bar{X} )</th>
<th>( (X - \bar{X})^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>68</td>
<td>-14</td>
<td>196</td>
</tr>
<tr>
<td>75</td>
<td>-7</td>
<td>49</td>
</tr>
<tr>
<td>77</td>
<td>-5</td>
<td>25</td>
</tr>
<tr>
<td>82</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>83</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>90</td>
<td>8</td>
<td>64</td>
</tr>
<tr>
<td>99</td>
<td>17</td>
<td>289</td>
</tr>
</tbody>
</table>

\[ \text{Range} = 99 - 68 = 31 \text{ points} \]

\[ \sigma^2 = \frac{\sum (X - \bar{X})^2}{m-1} = \frac{624}{7-1} = \frac{624}{6} = 104.0 \text{ PT}^2 \]

\[ \sigma = \sqrt{\frac{\sum (X - \bar{X})^2}{m-1}} = \sqrt{104} = 10.198 \ldots \]

Round to 10.2 PT

**Test 2**  \( \bar{X} = 82.0 \text{ points} \)

<table>
<thead>
<tr>
<th>( X )</th>
<th>( X - \bar{X} )</th>
<th>( (X - \bar{X})^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>73</td>
<td>-9</td>
<td>81</td>
</tr>
<tr>
<td>75</td>
<td>-7</td>
<td>49</td>
</tr>
<tr>
<td>81</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>82</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>85</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>86</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>92</td>
<td>10</td>
<td>100</td>
</tr>
</tbody>
</table>

\[ \text{Range} = 92 - 73 = 19 \text{ points} \]

\[ \sigma^2 = \frac{\sum (X - \bar{X})^2}{m-1} = \frac{256}{7-1} = \frac{256}{6} = 42.6 \text{ PT}^2 \]

\[ \sigma = \sqrt{\frac{\sum (X - \bar{X})^2}{m-1}} = \sqrt{42.6} = 6.53 \ldots \]

Round to 6.5 PT

Example: Use the coefficient of variation to compare the variation of test scores to the variation of study time for test 1 given

for scores:  \( \bar{X} = 82.0 \text{ points}, \ s = 10.2 \text{ points} \),

for study time:  \( \bar{X} = 11.4 \text{ hours}, \ s = 1.8 \text{ hours} \).

**For Scores**

\[ CVar = \frac{\sigma}{\bar{X}} \cdot 100\% = \frac{10.2 \text{ PT}}{82.0 \text{ PT}} \cdot 100\% \approx 12.4\% \]

**For Hours**

\[ CVar = \frac{\sigma}{\bar{X}} \cdot 100\% = \frac{1.8 \text{ hrs}}{11.4 \text{ hrs}} \cdot 100\% \approx 15.8\% \]
Finding the Sample Variance and Standard Deviation for Grouped Data

\[ s^2 = \frac{\sum f \cdot X_m^2 - \left( \frac{\sum f \cdot X_m}{n} \right)^2}{n-1} \]

Multiply each \( f \) by the corresponding \( X_m \) and add them all, then divide by \( n \).

3-3 Exercise 20  
Automotive Fuel Efficiency in MPG
Find the variance and standard deviation.

<table>
<thead>
<tr>
<th>Class</th>
<th>Frequency</th>
<th>( X_m )</th>
<th>( f \cdot X_m )</th>
<th>( f \cdot X_m^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.5 - 12.5</td>
<td>3</td>
<td>10</td>
<td>30</td>
<td>300</td>
</tr>
<tr>
<td>12.5 - 17.5</td>
<td>5</td>
<td>15</td>
<td>75</td>
<td>112.5</td>
</tr>
<tr>
<td>17.5 - 22.5</td>
<td>15</td>
<td>20</td>
<td>300</td>
<td>6000</td>
</tr>
<tr>
<td>22.5 - 27.5</td>
<td>5</td>
<td>25</td>
<td>125</td>
<td>312.5</td>
</tr>
<tr>
<td>27.5 - 32.5</td>
<td>2</td>
<td>30</td>
<td>60</td>
<td>1800</td>
</tr>
<tr>
<td>M = 30</td>
<td></td>
<td></td>
<td>( \sum (f \cdot X_m) )</td>
<td>( \sum f \cdot X_m^2 )</td>
</tr>
</tbody>
</table>

\[ \text{Variance} \quad s^2 = \frac{\sum f \cdot X_m^2 \cdot \frac{\sum f \cdot X_m}{m}^2}{m-1} \]

\[ = \frac{12350 - \left( \frac{590}{30} \right)^2}{30-1} \]

\[ = 25.747 \ldots \text{ Round to } 25.75 \text{ MPG}^2 \]

Since the raw data has one decimal place, the variance and standard deviation are rounded to two decimal places.

\[ \text{Standard Deviation} \quad \sigma = \sqrt{s^2} \approx 5.07 \text{ MPG} \]