Normal Distributions and the Empirical Rule

Normal Distribution – A data set that is characterized by the following criteria –

- The Mean and Median of the distribution are equal to the Mode.
- Most data values are clustered near the Mean (or Mode) so that the distribution has a well-defined peak.
- Data values are then spread evenly around the Mean (or Mode) so that the distribution is symmetric.
- Data values become increasingly rare as you move farther to the right and to the left of the Mean. This results in tapering tales on both ends of the curve.
- The variation is characterized by the standard deviation of the data distribution.

Note – This is sometimes also referred to as a “Normal Curve” or a “Bell-Shaped Curve.”

Empirical Rule - When a histogram of data is considered to meet the conditions of a “Normal Distribution”, (i.e. its graph is approximately bell-shaped), then it is often possible to categorize the data using the following guidelines... (Note: $\sigma$ → symbol used for standard deviation.)

- About 68% of the data (68.3%) is within one standard deviation ($\pm 1 \sigma$) of the mean ($\mu$).
- About 95% of the data (95.4%) is within two standard deviations ($\pm 2 \sigma$) of the mean ($\mu$).
- About 99.7% of the data (all or almost all) is within three standard deviations ($\pm 3 \sigma$) of the mean ($\mu$).

Note - This rule is also sometimes called the “68 – 95 – 99.7 Rule.”

The Empirical Rule is illustrated in the picture below.

Note: The Empirical Rule implies that a data set that is normally distributed has a width of approximately 6 standard deviations ($Width \approx 6 \sigma$).
**Standard Deviation** – A measure of how far data values are spread around the mean of a data set. It is computed as the square root of the variance. The actual formula for calculating the standard deviation depends on whether the data represents a population or is from a sample.

**Population Standard Deviation:**

\[
\sigma = \sqrt{\frac{\text{population variance}}{n}} = \sqrt{\frac{\sum (x_i - \mu)^2}{n}}
\]

**Sample Standard Deviation:**

\[
s = \sqrt{\frac{\text{sample variance}}{n-1}} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}
\]

**The “Range Rule of Thumb”** - The standard deviation is sometimes approximated by using the range of a data distribution according to the following ...

\[
\text{Standard Deviation} \approx \frac{\text{Range}}{4} = \frac{\text{High} - \text{Low}}{4}
\]

Note: This approach works well in data sets where the values are evenly distributed and there are not any outliers.

**Example** - For the following data set, approximate the standard deviation using the range rule of thumb.

<table>
<thead>
<tr>
<th>8.2</th>
<th>8.8</th>
<th>9.2</th>
<th>10.6</th>
<th>12.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.4</td>
<td>9.0</td>
<td>9.7</td>
<td>11.6</td>
<td>14.0</td>
</tr>
<tr>
<td>8.5</td>
<td>9.2</td>
<td>10.4</td>
<td>11.8</td>
<td>15.9</td>
</tr>
<tr>
<td>8.8</td>
<td>9.2</td>
<td>10.5</td>
<td>12.6</td>
<td>16.1</td>
</tr>
</tbody>
</table>

\[
\text{Lowest data point} = 8.2 \quad \text{and} \quad \text{Highest data point} = 16.1
\]

\[
\sigma = \frac{16.1 - 8.2}{4} = 1.975
\]

**Example** - If for a certain data set, the standard deviation is \( \sigma = 4.5 \) and the mean is \( \mu = 20.2 \).

a. In between what two values is approx. 68% of the data? \( \Rightarrow \) Answer: \( \mu \pm 1\sigma = 20.2 \pm 4.5 \)

\[
= 15.7 \text{ and } 24.7
\]

b. In between what two values is approx. 95% of the data? \( \Rightarrow \) Answer: \( \mu \pm 2\sigma = 20.2 \pm 2(4.5) \)

\[
= 11.2 \text{ and } 29.2
\]

c. In between what two values is all or almost all of the data? \( \Rightarrow \) Answer: \( \mu \pm 3\sigma = 20.2 \pm 3(4.5) \)

\[
= 6.7 \text{ and } 33.7
\]