There are 2 methods in AP Stat that we use for assessing normality:

1. Use the **Empirical Rule** to see if approximately 68% of all data points are within $\mu \pm 1\sigma$, 95% of all data points are within $\mu \pm 2\sigma$, and 99.7% of all data points are within $\mu \pm 3\sigma$. In general, this rule works well for large data sets, but not so well for small sets.

2. Use the **normal probability plot**. The interpretation of this plot is that the more linear the points in the plot, the more normal the original data.

The TI-83 will graph the normal probability plot for you, but here is an example of how it is constructed.

**Example**: Determine if the following data are normally distributed:

32.1  24.9  29.4  40.2  36.2  33.0  28.5  33.4  38.1

You must first determine the percentile for each data point. The formula is shown at the right.

The smallest observation is 24.9. It is the first of 9 scores, so the percentile associated with it would be:

$$\text{Percentile} = \left( \frac{(0 + 0.5)}{9} \right) \times 100 = 5.56$$

This means that 5.56% of the area is to the left of our value. Then we find the z-score associated with this percentile (either using the table or using the invNorm(.0556) function on the calculator). The z-score is around $-1.59$. Therefore, the first point on our graph will be $(24.9, -1.59)$. The values for the other eight points are found similarly.

Complete the table at the right.

<table>
<thead>
<tr>
<th>Data (ascending order)</th>
<th>Position</th>
<th>Percentile</th>
<th>z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>24.9</td>
<td>1</td>
<td>$(0 + 0.5) / 9 \times 100 = 5.56$</td>
<td>$-1.59$</td>
</tr>
<tr>
<td>28.5</td>
<td>2</td>
<td>16.67</td>
<td>$-0.97$</td>
</tr>
<tr>
<td>29.4</td>
<td>3</td>
<td>27.78</td>
<td>$-0.59$</td>
</tr>
<tr>
<td>32.1</td>
<td>4</td>
<td>38.89</td>
<td>$-0.28$</td>
</tr>
<tr>
<td>33.0</td>
<td>5</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>33.4</td>
<td>6</td>
<td>61.11</td>
<td>0.28</td>
</tr>
<tr>
<td>36.2</td>
<td>7</td>
<td>72.22</td>
<td>0.59</td>
</tr>
<tr>
<td>38.1</td>
<td>8</td>
<td>83.33</td>
<td>0.97</td>
</tr>
<tr>
<td>40.2</td>
<td>9</td>
<td>94.44</td>
<td>1.59</td>
</tr>
</tbody>
</table>

Graphing is done by plotting (data value, z-score) as the $(x, y)$ coordinates.

Graph the data from above.
Below are Henry Cavendish’s 29 measurements of the density of the earth, made in 1978. The data gives the density of the earth as a multiple of the density of water.

5.50  5.61  4.88  5.07  5.26  5.55  5.36  5.29  5.58  5.65  
5.57  5.53  5.62  5.29  5.44  5.34  5.79  5.10  5.27  5.39  
5.42  5.47  5.63  5.34  5.46  5.30  5.75  5.68  5.85  

1) Sketch a stemplot and boxplot. Is it fairly symmetric or skewed in either direction?

Slightly skewed left, but fairly symmetric between Q1 and Q3.

2) Check how closely they follow the Empirical (68-95-99.7) rule. Find \( \mu \) and \( \sigma \) and create a histogram. Compare the percents of the observations in each class with the Empirical rule.

\[
\begin{align*}
\mu &= 5.45 \\
\sigma &= .22 \\
\mu - \sigma &= 5.23 & \mu - 2\sigma &= 5.01 & \mu - 3\sigma &= 4.79 & \mu - 4\sigma &= 4.57 \\
\mu + \sigma &= 5.67 & \mu + 2\sigma &= 5.89 & \mu + 3\sigma &= 6.11 & \mu + 4\sigma &= 6.33
\end{align*}
\]

within 1\( \sigma \): 75.86% should be 68%

within 2\( \sigma \): 96.55% should be 95%

within 3\( \sigma \): 100% should be 99.7%

To make a histogram I set the window to go from \( \mu - 4\sigma \) to \( \mu + 4\sigma \) (or 4.57 to 6.33) with a scale of \( \sigma \) (or .22).

3) Construct a Normal probability plot. Does this tell us it is Normal or not Normal?

The Normal probability plot is approximately linear, so the data is approximately Normal.