Comparing Values from Different Data Sets

The standard deviation is useful when comparing data values that come from different data sets. If the data sets have different means and standard deviations, it can be misleading to compare the data values directly.

- For each data value, calculate how many standard deviations the value is away from its mean.
- Use the formula: \( \text{value} = \text{mean} + (\text{#ofSTDEVs})(\text{standard deviation}) \); solve for \( \text{#ofSTDEVs} \).
- \( \text{#ofSTDEVs} = (\text{value} - \text{mean}) / \text{standard deviation} \)
- Compare the results of this calculation.

\#ofSTDEVs is often called a "z-score"; we can use the symbol \( z \). In symbols, the formulas become:

\[
\begin{array}{|c|c|c|}
\hline
\text{Sample} & x = \bar{x} + z s & z = \frac{x - \bar{x}}{s} \\
\hline
\text{Population} & x = \mu + z \sigma & z = \frac{x - \mu}{\sigma} \\
\hline
\end{array}
\]

TABLE 3
EXAMPLE 3

Two students, John and Ali, from different high schools, wanted to find out who had the highest G.P.A. when compared to his school. Which student had the highest G.P.A. when compared to his school?

SOLUTION

For each student, determine how many standard deviations (#ofSTDEVs) his GPA is away from the average, for his school.

For John, \( z = \text{#ofSTDEVs} = (2.85 - 3.0)/0.7 = -0.21 \)

For Ali, \( z = \text{#ofSTDEVs} = (77 - 80)/10 = -0.3 \)

John has the better G.P.A. when compared to his school because his G.P.A. is 0.21 standard deviations below his school's mean while Ali's G.P.A. is 0.3 standard deviations below his school's mean.

John's z-score of −0.21 is higher than Ali's z-score of −0.3. For GPA, higher values are better, so we conclude that John has the better GPA when compared to his school.
The following lists give a few facts that provide a little more insight into what the standard deviation tells us about the distribution of the data.

For ANY data set, no matter what the distribution of the data is:
- At least 75% of the data is within 2 standard deviations of the mean.
- At least 89% of the data is within 3 standard deviations of the mean.
- At least 95% of the data is within 4 1/2 standard deviations of the mean.
- This is known as Chebyshev's Rule.

For data having a distribution that is MOUND-SHAPED and SYMMETRIC:
- Approximately 68% of the data is within 1 standard deviation of the mean.
- Approximately 95% of the data is within 2 standard deviations of the mean.
- More than 99% of the data is within 3 standard deviations of the mean.
- This is known as the Empirical Rule.

It is important to note that this rule only applies when the shape of the distribution of the data is mound-shaped and symmetric. We will learn more about this when studying the "Normal" or "Gaussian" probability distribution in later chapters.