Chebyshev’s Rule

For any distribution no matter how far from Bell shaped it is
the proportion (or fraction) of the data that
lies within $K$ standard deviations of the mean
is at least $1 - \frac{1}{K^2}$ for all $K > 1$

What is the difference between
The Empirical Rule for Data Under a Bell Shaped Curve
and Chebyshev’s Rule

The Empirical Rule requires that the data must be bell shaped. This requires the data to be very “well behaved”. This means the graph of the data forms a symmetric curve with the mean centered in the middle.

Chebyshev’s Rule is true for any data set. The graph of the data can be skewed left, skewed right, bi-modal or any other type of graph that is not normal.

If there is no requirement for the graph to be a certain shape it is harder to guarantee a given percent of the data falls within a given standard deviation. Requiring the data to be bell shaped allows a higher degree of certainty about the distribution of the data. Chebyshev’s Rule allows us to state that at least some % of the data set that falls within $K$ standard deviations of the mean in a “worst case” distribution.

Example 1

Chebyshev’s Rule says at least 75% of all the data in a data set falls within 2 standard deviation of the mean no matter how skewed the graph of the data is. This means for ANY DATA SET, 75% of all the data in a data set falls within 2 standard deviations of the mean. If the data set is bell shaped then the Empirical Rule allows us to be even more precise and say that at least 95% of all the data in a bell shaped data set falls within 2 standard deviations of the mean.

Example 2

Chebyshev’s Rule says at least 89% of all the data in a data set falls within 3 standard deviation of the mean no matter how skewed the graph of the data is. This means for ANY DATA SET, 89% of all the data in a data set falls within 3 standard deviations of the mean. If the data set is bell shaped then the Empirical Rule allows us to be even more precise and say that at least 99.7% of all the data in a bell shaped data set falls within 3 standard deviations of the mean.
Example 1

1A) What percent of the data falls within 2 standard deviations of the mean.

\[ K = 2 \text{ so} \]
\[ 1 - \frac{1}{K^2} = 1 - \frac{1}{2^2} \]
\[ = 1 - \frac{1}{4} = \frac{3}{4} \]
\[ = .75 \]
\[ = \text{at least 75\% of all the data falls within 2 SD of the mean} \]

1B) What percent of the data falls within 3 standard deviations of the mean.

\[ K = 3 \text{ so} \]
\[ 1 - \frac{1}{K^2} = 1 - \frac{1}{3^2} \]
\[ = 1 - \frac{1}{9} = \frac{8}{9} \]
\[ = .89 \]
\[ = 89\% \text{ of all the data falls within 3 SD of the mean} \]

1C) What percent of the data falls within 1.7 standard deviations of the mean.

\[ K = 1.7 \text{ so} \]
\[ 1 - \frac{1}{K^2} = 1 - \frac{1}{1.7^2} \]
\[ = .65 \]
\[ = \text{at least 65\% of all the data falls within 1.7 SD of the mean} \]
Finding $K$ given a point in the data set and the mean and standard deviation of the data

A data point $x$ is $K$ standard deviations from the mean where

$$K = \frac{|x - \text{the mean}|}{\text{the standard deviation}}$$

**Example 2**

A non normal set of data has a mean of 30 and a standard deviation of 6.

How many standard deviations is 42 from the mean?

2A) $x = 42$

Set up: $K = \frac{|42 - 30|}{6} = 2$

Interpretation: the data point 42 is 2 SD from the mean.

How many standard deviations is 28 from the mean?

2B) $x = 26$

Set up: $K = \frac{|26 - 30|}{6} = 2$

Interpretation: the data point 26 is 2 SD from the mean.

**Example 3**

A non normal set of data has a mean of 52 and a standard deviation of 6.

How many standard deviations is 70 from the mean?

3A) $x = 70$

Set up: $K = \frac{|70 - 52|}{6} = 3$

Interpretation: the data point 70 is 3 SD from the mean.

How many standard deviations is 34 from the mean?

3B) $x = 34$

Set up: $K = \frac{|34 - 52|}{6} = 3$

Interpretation: the data point 34 is 3 SD from the mean.
Example 4

The average number of cars in parking lot lot D at 6 pm is 30 cars with a standard deviation of 5 cars. The data is skewed (it IS NOT bell shaped).

Use Chebyshev’s Rule to state what percent of the data falls within 20 to 40 cars.

If $x = 40$ then $K = \frac{40 - 30}{5} = 2$

If $x = 20$ then $K = \frac{20 - 30}{5} = 2$

40 is 10 units from 30. if the standard deviation is 5 then 40 is 2 standard deviations from 30.

20 is 10 units from 30. if the standard deviation is 5 then 20 is 2 standard deviations from 30

Since both $x$ are 2 standard deviations from the mean then $k = 2$

$K = 2$ so

$1 - \frac{1}{K^2} = 1 - \frac{1}{2^2}$

$= 1 - \frac{1}{4} = \frac{3}{4}$

$= .75$

= at least 75% of all the data falls within 2 SD of the mean

Example 5

The average number of people in the PLE at six pm is 100 people with a standard deviation of 15 people. The data is bimodal. (it IS NOT bell shaped).

Use Chebyshev’s Rule to state what percent of the data falls within 40 to 160 people.

If $x = 160$ then $K = \frac{160 - 100}{15} = 4$

If $x = 40$ then $K = \frac{160 - 100}{15} = 4$

Since both $x$ are 4 standard deviations from the mean then $k = 4$

$K = 4$ so

$1 - \frac{1}{K^2} = 1 - \frac{1}{4^2}$

$= 1 - \frac{1}{16} = \frac{15}{16}$

$= .94$

= at least 94% of all the data falls within 4 SD of the mean