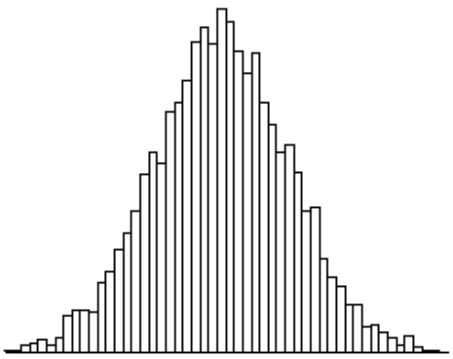
Standard Deviation Rule

Learning Objective: Apply the standard deviation rule to the special case of distributions having the "normal" shape.

The Standard Deviation Rule

In the previous activity we tried to help you develop better intuition about the concept of standard deviation. The rule that we are about to present, called "The Standard Deviation Rule" (also known as "The Empirical Rule") will hopefully also contribute to building your intuition about this concept.

Consider a symmetric mound-shaped distribution:



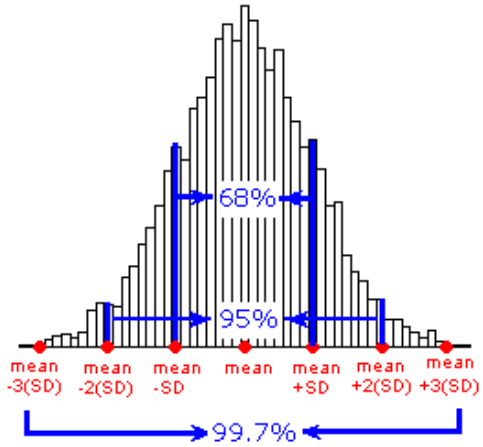
For distributions having this shape (also known as the **normal** shape), the following rule applies:

The Standard Deviation Rule:

- Approximately 68% of the observations fall within 1 standard deviation of the mean.
- Approximately 95% of the observations fall within 2 standard deviations of the mean.

- Approximately 99.7% (or virtually all) of the observations fall within 3 standard deviations of the mean.

The following picture illustrates this rule:

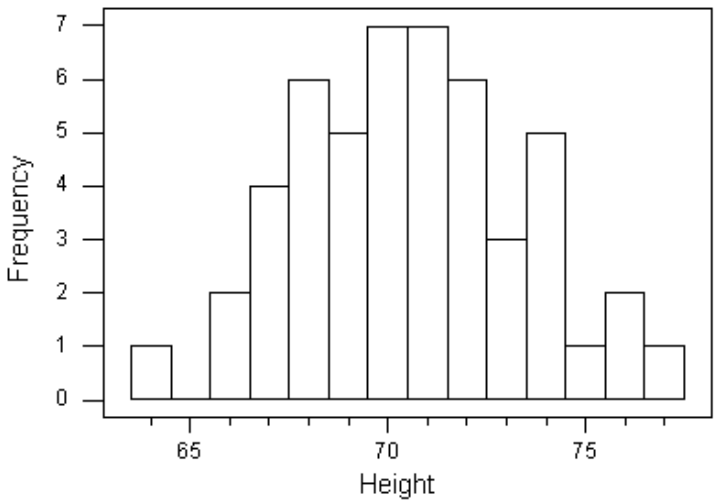


This rule provides another way to interpret the standard deviation of a distribution, and thus also provides a bit more intuition about it.

To see how this rule works in practice, consider the following example:

Example: Male Height

The following histogram represents height (in inches) of 50 males. Note that the data are roughly normal, so we would like to see how the Standard Deviation Rule works for this example.



Below are the actual data, and the numerical summaries of the distribution. Note that the key players here, the mean and standard deviation, have been highlighted.

64	[66	66	67	67	67	67	[68	68	68
68	68	68	69	69	69	69	69	70	70
70	70	70	70	70	71	71	71	71	71
71	71	72	72	72	72	72	72	73	73
73]	74	74	74	74	74	75	76	76]	77

Statistic	Height
N	50
Mean	70.58
StDev	2.858
min	64
Q1	68
Median	70.5
Q3	72
Max	77

To see how well the Standard Deviation Rule works for this case, we will find what percentage of the observations falls within 1, 2, and 3 standard deviations from the mean, and compare it to what the Standard Deviation Rule tells us this percentage should be.

Interval	Mean-SD, Mean+SD (67.7 , 73.4)	Mean-2(SD),Mean+2(SD) (64.9 , 76.3)	Mean-3(SD), Mean+3(SD) (62 , 79.2)
Percentage of Observations in interval	34 observations 34/50 = 68%	48 observations 48/50 = 96%	All 50 observations 50/50 = 100%
SD Rule says...	68%	95%	99.7%

It turns out the Standard Deviation Rule works **very well** in this example.
