

Standard Deviations, Normal Distributions, and Percentiles

Standard Deviation

A standard deviation is a summary measure of the distribution of a quantitative variable. Its value is an attempt to quantify how spread out the values are from average/center.

The standard deviation is a typical distance between one data value and the average of all data. Typically, some data values will be about one standard deviation away from average. Many values will be closer than one standard deviation from average, and some values will be farther than one standard deviation from average.

* In many distributions, a majority of values (maybe 60%, 70%, or 80%) will be within **one** standard deviation of average. In many distributions, almost all values (90% to 100%) will be within **two** standard deviations of average.

Normal distribution

A normal distribution occurs when there are so many individuals in a population that a histogram becomes a smooth curve, bell-shaped and symmetrical. Many populations can be closely approximated by a normal distribution.

* In a perfect normal distribution, 68% of values (actually 68.26%) are within one standard deviation of the mean. Also, 95% of values (actually 95.44%) are within two standard deviations of the mean.

Percentiles

A percentile tells you that a given value is greater than a certain percent of the values for the population.

* If you have a normal distribution, percentiles are related to z -scores and the areas found in the normal distribution table. For example, the 20-percentile value is found by looking up 0.2000 in the middle of the table, giving a z -score of approximately -0.84 . The data value can then be found using the mean, μ , the standard deviation, σ , and some algebra.

* If your distribution is not normally distributed, percentiles are **not** found using z -scores and the table; *this table* is for normal distributions only. Instead, you must sort the list and then look at the value a certain percent of the way through the list. For example, the 20-percentile value is the value that occurs 20% of the way through the sorted list.