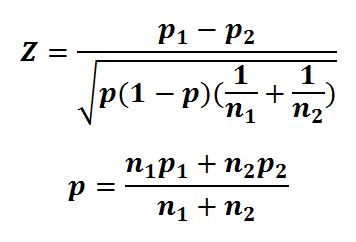


The mean and standard deviation are two fundamental statistical measures with distinct yet complementary properties. The **mean**, or average, represents the central tendency of a dataset, providing a single value that summarizes all the numbers in the set. It is sensitive to every value in the data, meaning that even a single extremely large or small value (an outlier) can significantly affect the mean. On the other hand, the **standard deviation** measures the amount of variation or dispersion in the dataset. A low standard deviation indicates that the data points are closely clustered around the mean, while a high standard deviation suggests that the values are spread out over a wider range. Both the mean and standard deviation are widely used in statistics to describe and compare datasets, and they are especially important in fields such as science, finance, and engineering for understanding patterns, consistency, and reliability in data. 

The **z-test** and **p-value** are important concepts in statistical hypothesis testing, each with distinct properties. The z-test is used to determine whether there is a significant difference between sample and population means, or between two sample means, especially when the sample size is large and the population variance is known. It assumes the data are normally distributed and calculates a z-score, which measures how many standard deviations a data point is from the mean. The p-value, on the other hand, is not a test itself but a probability that quantifies the strength of evidence against the null hypothesis. It represents the likelihood of obtaining a result at least as extreme as the observed one, assuming the null hypothesis is true. A small p-value suggests strong evidence against the null hypothesis, while a large p-value indicates weak evidence. Together, the z-test provides the test statistic and the p-value helps decide whether to reject or fail to reject the null hypothesis, making both essential tools for drawing conclusions from data in scientific research and data analysis.



The **67-95-99 rule** (more commonly known as the **68-95-99.7 rule** or the **empirical rule**) describes how data is distributed in a normal (Gaussian) distribution. According to this rule, about 68% of the data falls within one standard deviation of the mean, about 95% falls within two standard deviations, and about 99.7% falls within three standard deviations. This means that for a bell-shaped Gaussian curve, the vast majority of values are clustered close to the mean, and very few are found far from it. The rule is widely used in statistics to quickly estimate the spread of data and to identify outliers or unusual observations. It only applies to data that is normally distributed and helps in understanding probabilities, making predictions, and setting thresholds in quality control and risk analysis





A **log-normal distribution** is a continuous probability distribution in which the logarithm of the random variable is normally distributed. In other words, if a variable X*X* is log-normally distributed, then Y=ln⁡(X)*Y*=ln(*X*) follows a normal (Gaussian) distribution. This means that while the normal distribution can take both positive and negative values, the log-normal distribution only takes positive real values, making it especially useful for modeling data that cannot be negative and often displays a right-skewed, long-tailed pattern.

The log-normal distribution commonly arises in situations where a quantity results from the multiplicative product of many independent, positive random variables—such as in modeling incomes, biological measurements, stock prices, and certain natural phenomena. Its probability density function is defined for x>0 and is characterized by two parameters: the mean (μ*μ*) and standard deviation (σ*σ*) of the variable’s natural logarithm, not of the variable itself.

