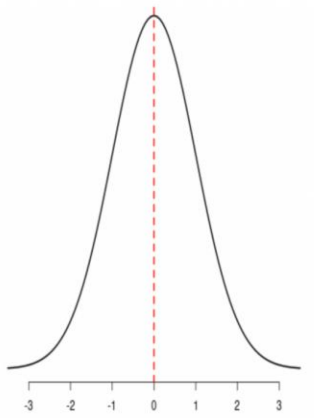
ASSIGNMENT 3

1) What does symmetric distribution mean?

In statistics, a symmetric distribution is a distribution in which the left and right sides mirror each other. The most well-known symmetric distribution is the normal distribution, which has a distinct bell-shape. If you were to draw a line down the center of the distribution, the left and right sides of the distribution would perfectly mirror each other:



In statistics, skewness is a way to describe the symmetry of a distribution. This value can be negative, zero, or positive. For symmetric distributions, the skewness is zero. In a symmetrical distribution, the mean, median, and mode are all equal. Mean: The average value, Median: The middle value, Mode: The value that occurs most often. In a symmetrical distribution, each of these values is equal to each other.

2) What is left skewed distribution and right skewed distribution?

Skewness is a way to describe the symmetry of a distribution.

A distribution is left skewed if it has a “tail” on the left side of the distribution. Left skewed distributions are sometimes called “negatively-skewed” distributions. The skewness value of any distribution showing a negative skew is always less than zero. Left Skewed Distribution: Mean < Median < Mode.

A distribution is right skewed if it has a “tail” on the right side of the distribution. Right skewed distributions are sometimes called “positively-skewed” distributions. A positively skewed distribution assumes a skewness value of more than zero. Right Skewed Distribution: Mode < Median < Mean.

3) Where are long-tailed distributions used?

A long-tailed distribution is a probability distribution that has a large number of occurrences that are far from the central part of the distribution. It is a type of distribution where the tail drops off gradually toward the end of the curve. The Pareto principle and the product sales distribution are good examples to denote the use of long-tailed distributions. Also, it is widely used in classification and regression problems. Long-tail distributions is the basis for many business models.

4) What is the central limit theorem?

In probability theory, the central limit theorem (CLT) states that the distribution of a sample variable approximates a normal distribution (i.e., a "bell curve") as the sample size becomes larger, assuming that all samples are identical in size, and regardless of the population's actual distribution shape.

CLT is a statistical premise that, given a sufficiently large sample size from a population with a finite level of variance, the mean of all sampled variables from the same population will be approximately equal to the mean of the whole population. Furthermore, these samples approximate a normal distribution, with their variances being approximately equal to the variance of the population as the sample size gets larger, according to the law of large numbers.

A key aspect of CLT is that the average of the sample means and standard deviations will equal the population mean and standard deviation. A sufficiently large sample size can predict the characteristics of a population more accurately.

5) What are observational and experimental data in statistics?

Observational data correlates to the data that is obtained from observational studies, where variables are observed to see if there is any correlation between them. There is no human intervention. The researcher does not influence the population in any way.

Experimental data is derived from experimental studies, where certain variables are held constant to see if any discrepancy is raised in the working. Researcher manipulates the variable and tries to determine how the manipulation influences the other variable. It is a data that is collected from an experimental study that involves taking measurements which can be manipulated.