### **What is the law of large numbers, and how does it relate to probability theory?**

The law of large numbers (LLN) in probability states that as the sample size increases, the sample mean tends to approach the population mean. When we repeat an experiment multiple times and average the results, we get a value that is close to the expected value.

Let's understand it with the help of an example:

If we flip a coin, there is a 50% chance of getting a head or a tail. So, 50% is the expected value (population mean)

If we toss a coin 4 times, we get 1 time head, so the probability of a head is 25%

Now we toss a coin 100 times, we get 47 times heads, so the probability of heads is 47%

Now again tossing 1000 times, we get 543 heads, so the probability is 54.3%

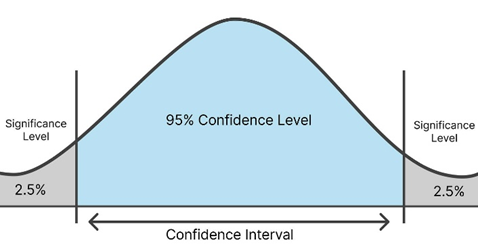
Here we are increasing our sample size (number of trials), which is tossing of coin, we tend to approach 50%, which is the population mean.

We can conclude that it forms the basis for the statistical inference that allows us to make predictions and estimate parameters based on sample data. The LLN illustrates the principle that larger samples provide more accurate estimates of population parameters.

### **How is a confidence interval defined in statistics?**

A confidence interval is the range of values within which we expect a particular population parameter, like the mean, to fall; it’s a way to express the uncertainty around the estimate obtained from a sample of data.

Confidence level is usually expressed as a percentage that indicates how sure we are that the true value lies within the interval.



Let’s understand with an example:

If a YouTube channel has a number of subscribers and the average age of them lies between 25 to 32, then this is a confidence interval, and the confidence level is 95 %, which means that we are 95 % confident that the average age is between 25 to 32.

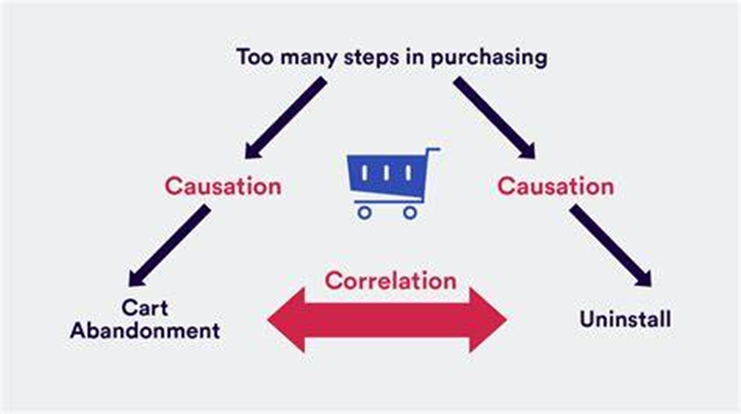
If we calculate a 95% confidence interval for a population’s average height, and we randomly select a sample of 50 students and calculate the average height to be 165 cm for instance, and the result is a range of 160 to 170 cm, this suggests that if we were to take multiple samples and create confidence intervals in the same manner, we should anticipate that approximately 95% of those intervals would contain the population’s true average height.

### **What is the difference between correlation and causation?**

Correlation is the relationship between variables. There are positive, negative, and zero correlations. If one value goes up, then the other also goes up, or vice versa, then it is a positive correlation. If one value goes higher and the other goes lower, then it is a negative correlation. If no relation exists, then there is zero correlation.

Causation is where one variable directly influences change in another variable. There is an effect relationship between variables.

Correlation does not imply causation.If a correlation exists between variables, it does not mean that one variable has caused the change in the other variable.



Example:

Data is collected on sales, and it is found that there is a positive correlation between the sales of ice cream and the sales of sunscreen. If we try to plot a scatter plot to visualize the relationship, we can see that as the sale of ice cream increases, the sale of sunscreen increases. It does not imply that an increase in ice cream has caused people to buy sunscreen; there is no direct causal link between ice cream sales and sunscreen sales. There is a third variable that is summer weather, because of that, sales increase in both products.

### **What is the difference between Descriptive and Inferential Statistics?**

Descriptive and Inferential Statistics are types of Statistical analysis.

Descriptive statistics deals with the collection, organisation, analysis, interpretation, and presentation of data, it focuses on summarizing and describing the main features of the set of data.

Inferential Statistics deals with making conclusions and predictions about the population based on the sample.

Descriptive statistics uses methods like measures of central tendency, measures of dispersion, while inferential statistics uses techniques like probability, hypothesis testing, and ANOVA test to make predictions.

Example: Student’s marks in class. Descriptive statistics will give the information like average marks, range of marks, and most common marks of students, while the inferential statistics will predict the performance of students based on marks in future tests.

| S.No. | Descriptive Statistics | Inferential Statistics |
| --- | --- | --- |
| 1. | It gives information about raw data, which describes the data in some manner. | It makes inferences about the population using data drawn from the population. |
| 2. | It helps in organising, analyzing, and presenting data in a meaningful manner. | It allows us to compare data and make hypotheses and predictions. |
| 3. | It is used to describe a situation. | It is used to explain the chance of the occurrence of an event. |
| 4. | It explains already known data and is limited to a sample or population having a small size. | It attempts to reach a conclusion about the population. |
| 5. | It can be achieved with the help of charts, graphs, tables, etc. | It can be achieved by probability. |

### **Can two confidence intervals with different widths have the same confidence level?**

Yes, two confidence intervals with different widths can have the same confidence level. The width of the confidence interval is affected by sample size and variability. A wider interval means more uncertainty, while a narrow interval means greater precision.

Let’s understand with an example:

We are trying to catch fish using two different nets. Both nets have a 90% chance of catching fish, one net is wider and the other is narrower. So, the nets are of different sizes, but both give the same level of confidence to catch fish. If the net is wider, it will catch a large number of fish, and with that, it can catch large amounts of junk. If the net is narrow, it catches fewer fish but also less junk.