**Name:** Abhishek Srivastva

**Registration Number:** 19BCE10071

**Subject:** NAS1001 - NASCOM Future Skills - Data Analytics

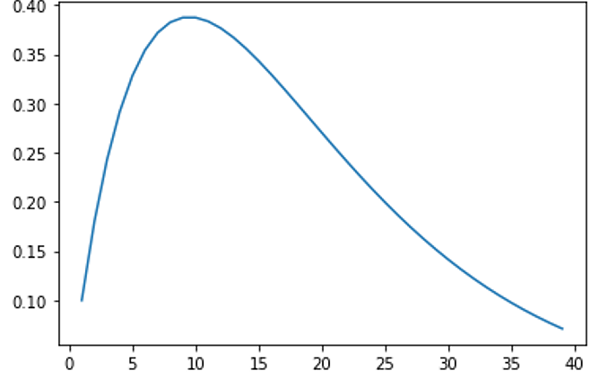
**Slot:**  A11+A12+A13

**Binomial Process**

A binomial process is a random counting system where there are n independent identical trials, each one of which has the same probability of success p, which produces s successes from those n trials (where 0 ≤ s ≤ n and n > 0 obviously).

A binomial distribution is a discrete probability distribution that has only two outcomes: success and failure.

The simplest example of a binomial process is the toss of a coin. If we define 'heads' as a success, each toss has the same probability of success p (0.5 for a fair coin). Then, for a given number of trials n (tosses of a coin), the number of successes will be s (the number of 'heads'. Each trial can be thought of as a random variable that returns either a 1 with probability p or a 0 with probability (1-p).



**Bernoulli Distribution**

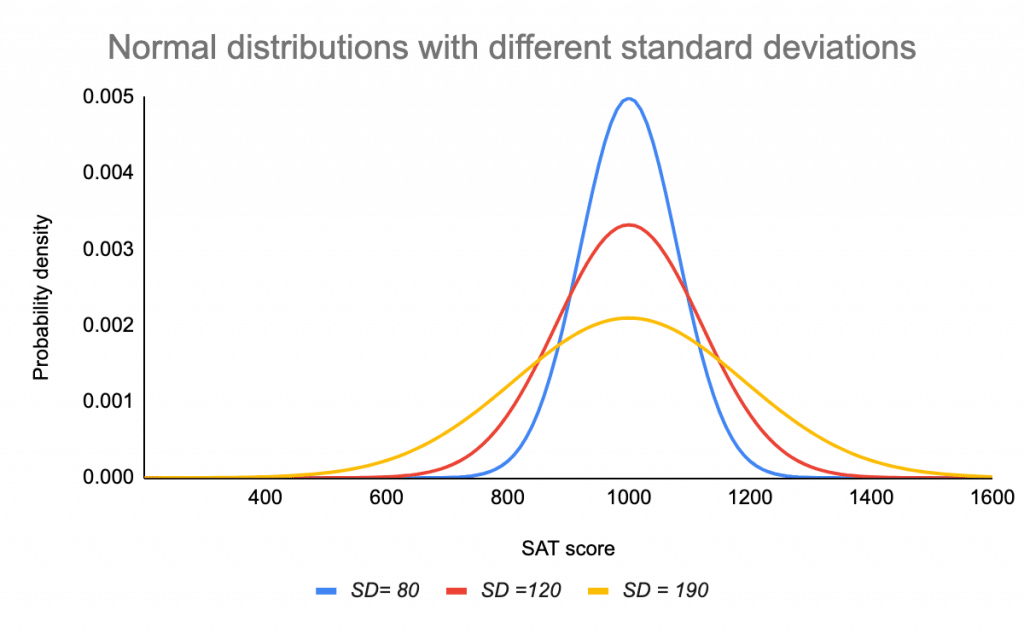
A Bernoulli distribution is a discrete probability distribution that can be used for any two outcomes, independent versions of the binomial distribution.

For example, the probability of getting a heads (a “success”) while flipping a coin is 0.5. The probability of “failure” is 1 – P (1 minus the probability of success, which also equals 0.5 for a coin toss).

It is a special case of the binomial distribution for n = 1. In other words, it is a binomial distribution with a single trial (e.g. a single coin toss).

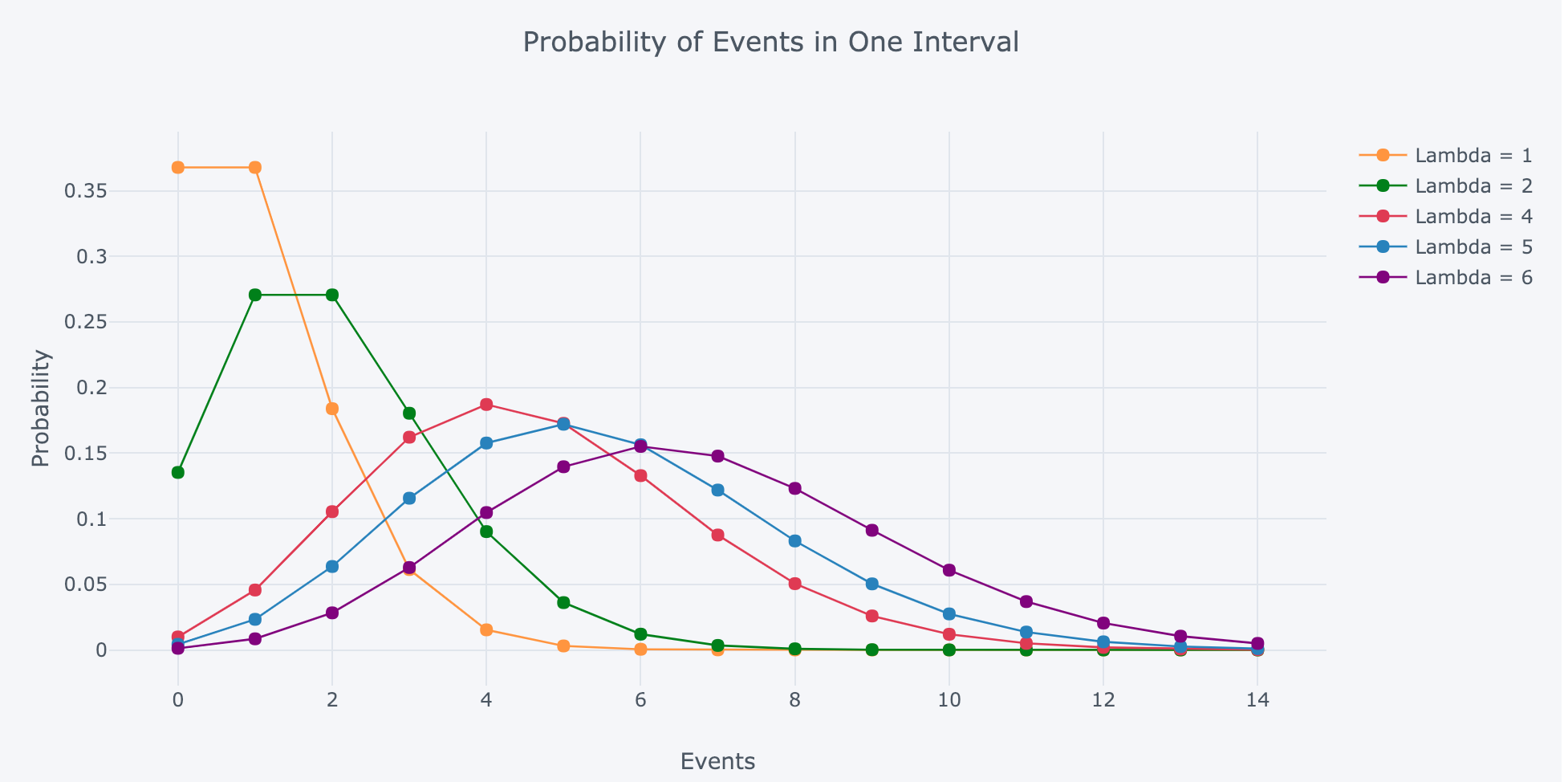
**Normal Distribution**

A normal distribution, sometimes called the bell curve, is a distribution that occurs naturally in many situations. The bell curve is symmetrical. Half of the data will fall to the left of the mean; half will fall to the right.

For example, the bell curve is seen in tests like the SAT and GRE. The bulk of students will score the average (C), while smaller numbers of students will score a B or D. An even smaller percentage of students score an F or an A.

**Poisson Process and Poisson Distribution**

In a Poisson Process - Events are independent of each other, the occurrence of one event does not affect the probability of another event, the average rate (events per time period) is constant and no two events cannot occur at the same time.

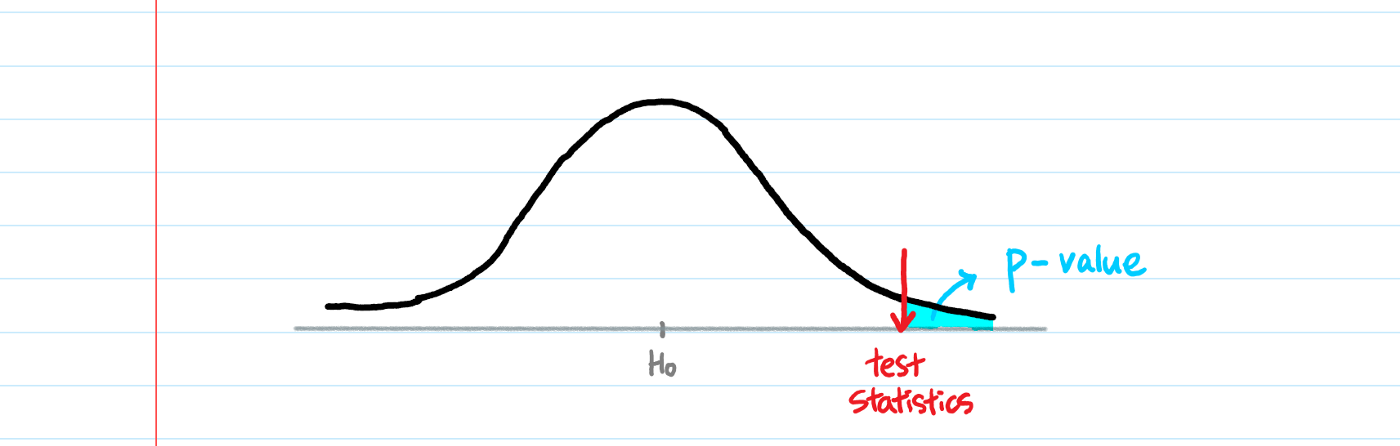


The Poisson distribution is the discrete probability distribution of the number of events occurring in a given time period, given the average number of times the event occurs over that time period.

The Poisson distribution is also commonly used to model financial count data where the tally is small and is often zero. As one example in finance, it can be used to model the number of trades that a typical investor will make in a given day, which can be 0 (often), or 1, or 2, etc.

**P Value**

A p value is used in hypothesis testing to help you support or reject the null hypothesis. The p value is the evidence against a null hypothesis. The smaller the p-value, the stronger the evidence that you should reject the null hypothesis.



For example, suppose a study comparing returns from two particular assets was undertaken by different researchers who used the same data but different significance levels. The researchers might come to opposite conclusions regarding whether the assets differ.

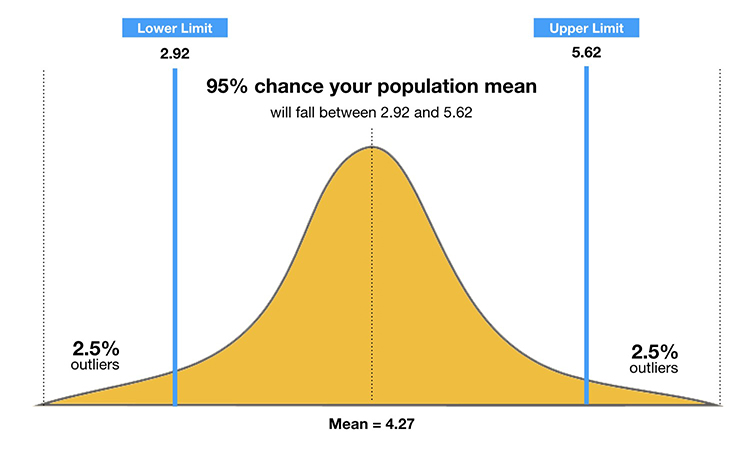
If one researcher used a confidence level of 90% and the other required a confidence level of 95% to reject the null hypothesis and the p-value of the observed difference between the two returns was 0.08 (corresponding to a confidence level of 92%), then the first researcher would find that the two assets have a difference that is statistically significant, while the second would find no statistically significant difference between the returns.

To avoid this problem, the researchers could report the p-value of the hypothesis test and allow the reader to interpret the statistical significance themselves. This is called a p-value approach to hypothesis testing.

**Confidence interval**

A confidence interval is how much uncertainty there is with any particular statistic. Confidence intervals are often used with a margin of error. It tells you how confident you can be that the results from a poll or survey reflect what you would expect to find if it were possible to survey the entire population.

Confidence intervals are your results and they are usually numbers. For example, you survey a group of pet owners to see how many cans of dog food they purchase a year. You test your statistic at the 99 percent confidence level and get a confidence interval of (200,300). That means you think they buy between 200 and 300 cans a year.

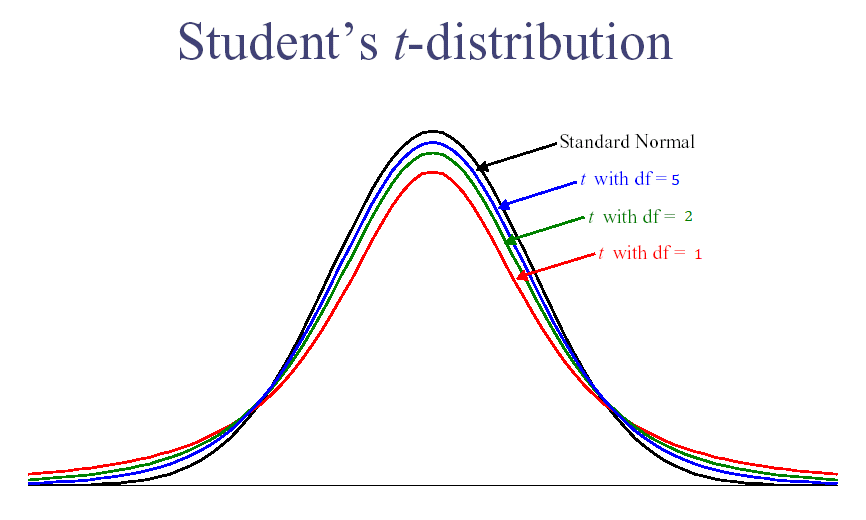
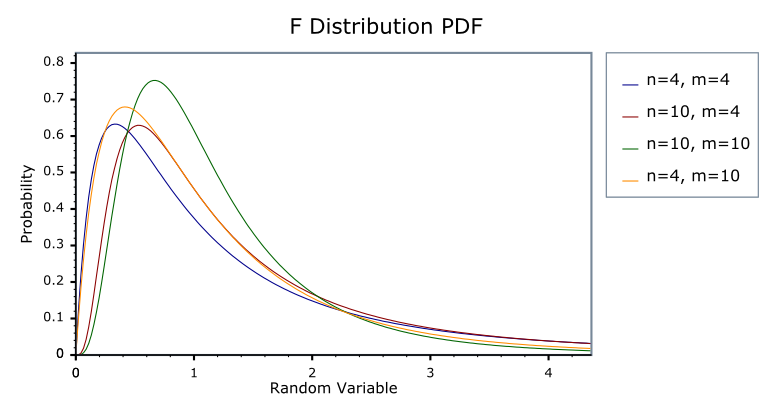


**T test**

A t-test is a statistical test that is used to compare the means of two groups. It is often used in hypothesis testing to determine whether a process or treatment actually has an effect on the population of interest, or whether two groups are different from one another. A t-test can only be used when comparing the means of two groups (a.k.a. pairwise comparison).

For example, a drug company may want to test a new cancer drug to find out if it improves life expectancy. In an experiment, there’s always a control group (a group who are given a placebo, or “sugar pill”). The control group may show an average life expectancy of +5 years, while the group taking the new drug might have a life expectancy of +6 years.

It would seem that the drug might work. But it could be due to a fluke. To test this, researchers would use a Student’s t-test to find out if the results are repeatable for an entire population.

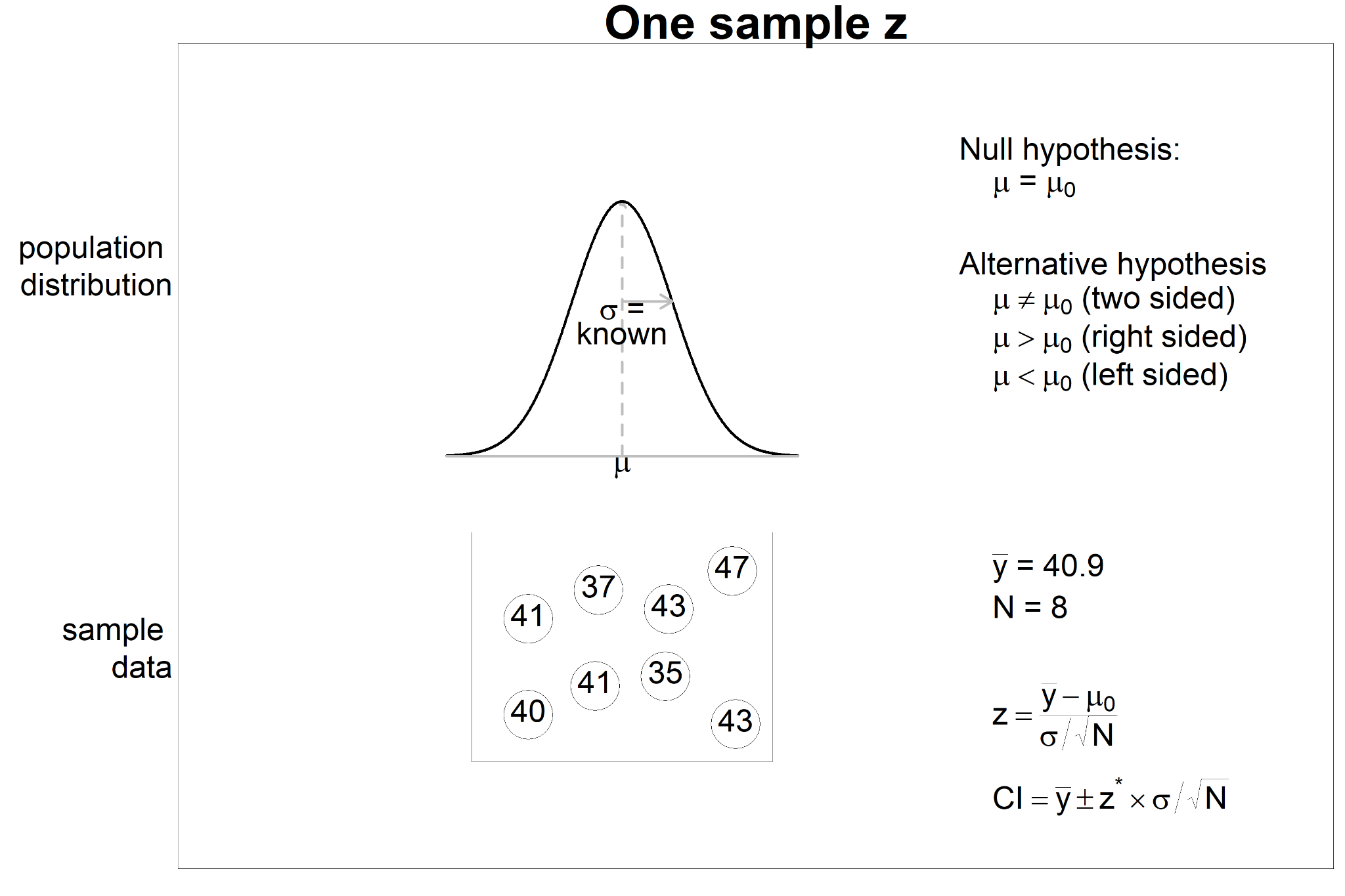


**Ftest**

F test calculates the ratio between the larger variance and the smaller variance. We use the F test when we want to check where the means of three or more groups are different or not. F-test is used to assess whether the variances of two populations (A and B) are equal.

**Z Test**

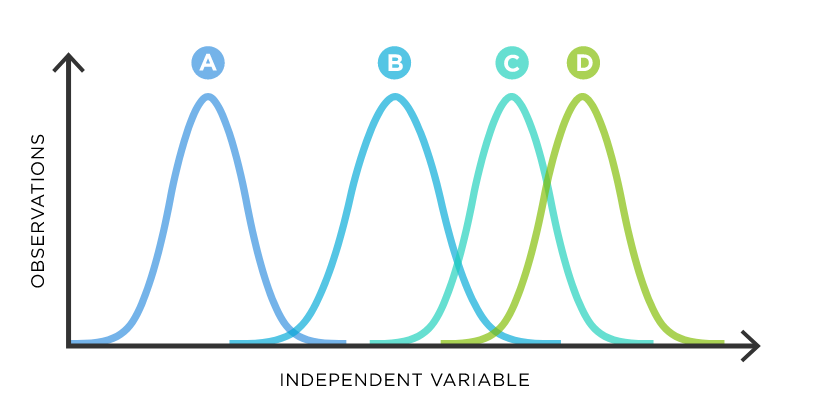
Z-test is a statistical method to determine whether the distribution of the test statistics can be approximated by a normal distribution. It is the method to determine whether two sample means are approximately the same or different when their variance is known and the sample size is large.



**ANOVA**

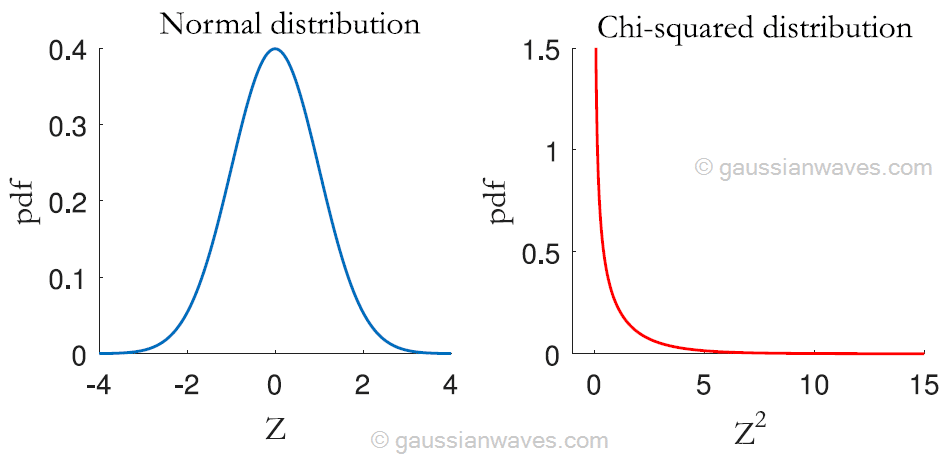
Analysis of variance (ANOVA) is a statistical technique that is used to check if the means of two or more groups are significantly different from each other. ANOVA checks the impact of one or more factors by comparing the means of different samples.

A common approach to figure out a reliable treatment method would be to analyse the days it took the patients to be cured. We can use a statistical technique which can compare these three treatment samples and depict how different these samples are from one another. We can use ANOVA to prove/disprove if all the medication treatments were equally effective or not.



**Chi Square**

A chi-square statistic is one way to show a relationship between two categorical variables. In statistics, there are two types of variables: numerical (countable) variables and non-numerical (categorical) variables. The chi-squared statistic is a single number that tells you how much difference exists between your observed counts and the counts you would expect if there were no relationship at all in the population.



**Interquartile range**

The interquartile range is a measure of where the “middle fifty” is in a data set. Where a range is a measure of where the beginning and end are in a set, an interquartile range is a measure of where the bulk of the values lie.

The Interquartile range formula measures the variability, based on dividing an ordered set of data into quartiles. Quartiles are three values or cuts that divide each respective part as the first, second, and third quartiles, denoted by Q1, Q2, and Q3, respectively.

