**Chi Square Formula**

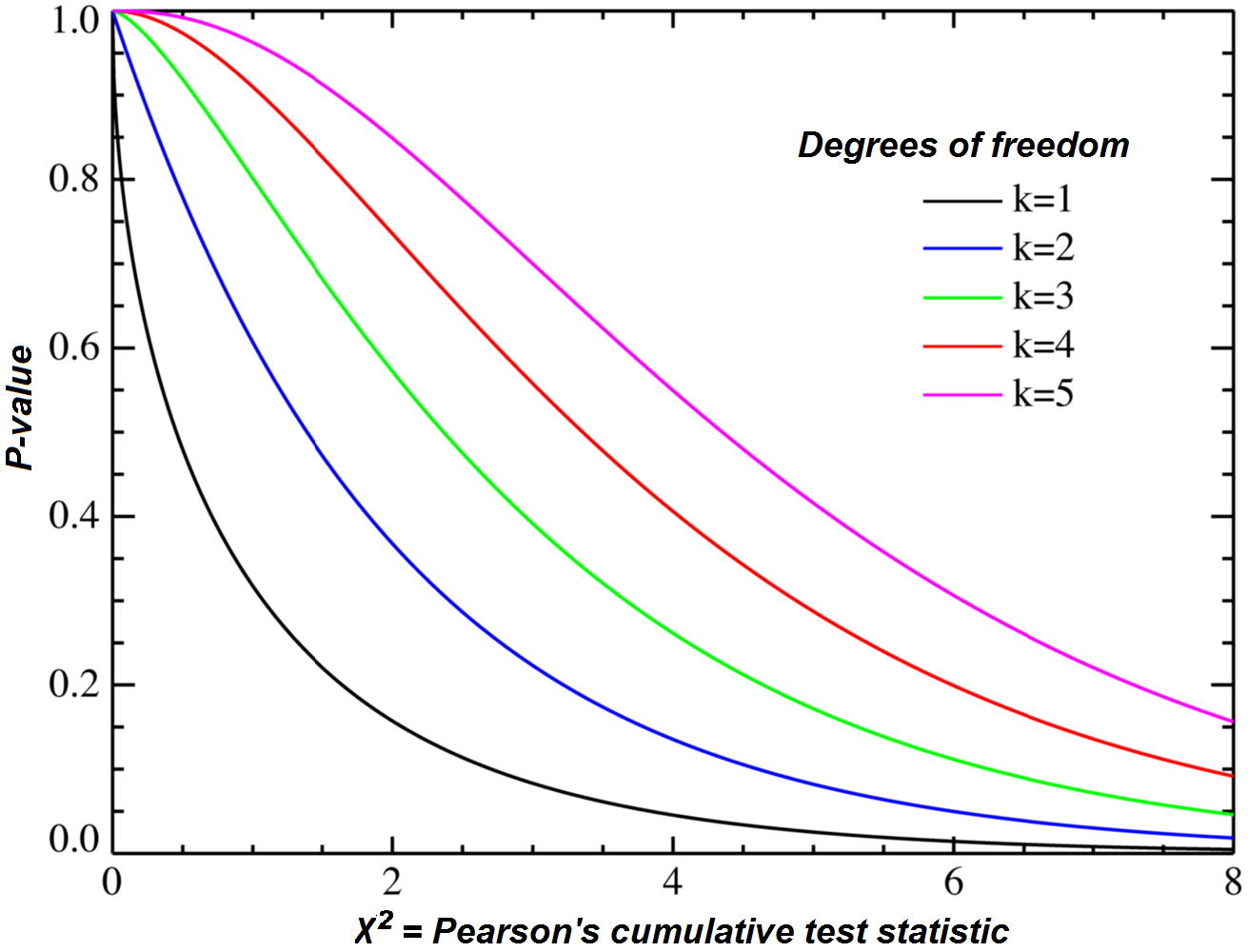
In statistics, various measurement methods are popularly used. For many experimental studies, we need a chi-square test to get conclusions. It is one of the most useful non-parametric statistics. The Chi Square test is used for data collection consist of probability distributed across various categories. And to know that whether the distribution is different from what would expect.

## **Chi Square Formula**

### **What is Chi-Square?**

Chi-square is a method that is used in statistics and it calculates the difference between observed and expected data values. It is used to find out how closely actual data fit with expected data. The value of chi-square will help us to get the answer to the question as to the significance of the difference in expected and observed data statistically. A small chi-square value will tell us that any differences in actual and expected data are due to some usual chance.

And hence the data is not statistically significant. Also, a large value will tell that the data is statistically significant and there is something causing the differences in data. From there, a statistician may explore factors that may be responsible for the differences.



Chi is a Greek symbol that looks like the letter x as we can see it in the formulas. To calculate the chi-square, we will take the square of the difference between the observed value O and expected value E values and further divide it by the expected value. Depending on the number of categories of the data, we end up with two or more values. Chi-square is the sum total of these values. However, the value that we are looking for is chi-square, we do not need to take the square root.

A very small Chi-Square test statistically means that the observed data fit in the expected data extremely well. A very large Chi-Square test statistically means that the data does not fit very well. If the chi-square value is very large, then we have to reject the null hypothesis.

Chi-Square is one way to show the relationship between two categorical variables. Generally, there are two types of variables in statistics such as numerical variables and non-numerical variables.

### **Formula for the Chi-Square Test**

The Chi-Square is denoted byχ2 and the formula is:

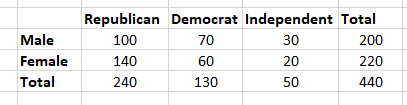
χ2=∑(O−E)2/E

Where,

* O: Observed frequency
* E:  expected frequency
* ∑:summation
* Chi 2:Chi Square Value

## **Example**

Let's say you want to know if gender has anything to do with political party preference. You poll 440 voters in a simple random sample to find out which political party they prefer. The results of the survey are shown in the table below:



To see if gender is linked to political party preference, perform a Chi-Square test of independence using the steps below.

### **Step 1: Define the Hypothesis**

H0: There is no link between gender and political party preference.

H1: There is a link between gender and political party preference.

### **Step 2: Calculate the Expected Values**

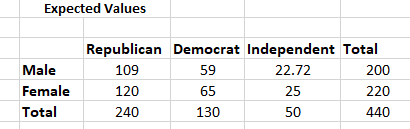
Now you will calculate the expected frequency.

Chi_Sq_formula_1.

For example, the expected value for Male Republicans is:

Chi_Sq_formula_2

Similarly, you can calculate the expected value for each of the cells.



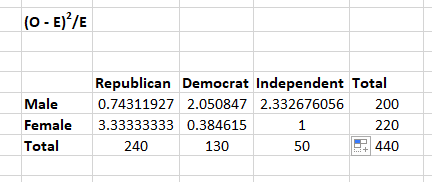
### **Step 3: Calculate (O-E)2 / E for Each Cell in the Table**

Now you will calculate the (O - E)2 / E for each cell in the table.

Where

O = Observed Value

E = Expected Value



### **Step 4: Calculate the Test Statistic X2**

X2 is the sum of all the values in the last table

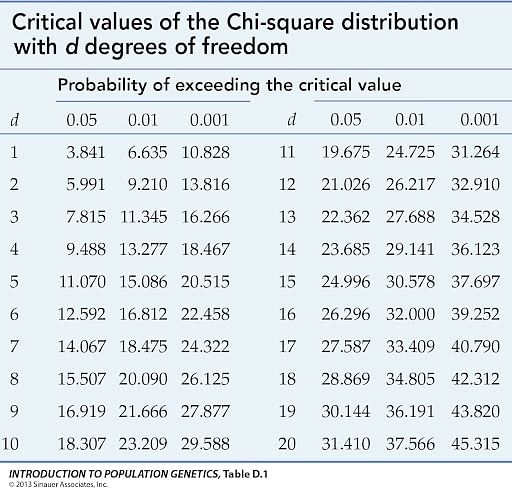
 =  0.743 + 2.05 + 2.33 + 3.33 + 0.384 + 1

 = 9.837

Before you can conclude, you must first determine the critical statistic, which requires determining our degrees of freedom. The degrees of freedom in this case are equal to the table's number of columns minus one multiplied by the table's number of rows minus one, or (r-1) (c-1). We have (3-1)(2-1) = 2.

Finally, you compare our obtained statistic to the critical statistic found in the chi-square table. As you can see, for an alpha level of 0.05 and two degrees of freedom, the critical statistic is 5.991, which is less than our obtained statistic of 9.83. You can reject our null hypothesis because the critical statistic is higher than your obtained statistic.

This means you have sufficient evidence to say that there is an association between gender and political party preference.

\*\*

## Types of T-tests

The t-test calculates a t-value, which is then compared to a critical value based on the size of the groups and the level of significance chosen by the researcher. If the calculated t-value is larger than the critical value, it means that there is a statistically significant difference between the two groups. If the calculated t-value is smaller than the critical value, it means that there is not enough evidence to suggest that the groups are different. There are three types of t-tests.

* If all of the groups come from one single population (like measuring before and after an experimental treatment), then we perform a **paired t-test**.
* If the groups under consideration come from two different populations (like two different species, or people from two separate cities), then we perform a **two-sample t-test**or **independent t-test**).
* If there is one group being compared against a standard value (like comparing the acidity of a liquid to a neutral pH of 7), then we perform a **one-sample t-test**.

## T-Test formula

The formula for the t-test (a.k.a. the student’s t test formula) is as follows –



* According to this formula, t is called the t-value,



* x1 and x2 are the means of the two groups which are being compared,
* S1 and S2 are the standard deviations of the first and second sets of values
* n1 and n2 are the numbers of observations of the first and second groups, respectively.

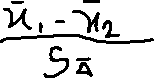
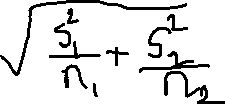
A greater t-value indicates that the difference between means is greater than the pooled standard error, which suggests a significant difference between the groups.

The calculated t-value through the t test formula can be compared against the values in a critical value chart to determine whether your t-value is greater than what would be expected by chance. If so, the null hypothesis can be rejected and it can be concluded that the two groups are different.

### T-test Example Problems with Solutions

**Example 1:**Is there a significant difference in test scores between 25 students who received in-person instruction and 25 students who received online instruction? The mean test score for the in-person group is 80 (SD = 5) and for the online group is 75 (SD = 7).

**Solution:**This is an independent samples t-test problem as the two groups being compared are independent of each other. To perform the t-test, we first calculate the t-value using the t test formula:



* where mean1 is the mean test score for the in-person group,
* mean2 is the mean test score for the online group
* s1 and s2 are the standard deviations for the two groups
* and n1 and n2 are the sample sizes.

Substituting the numbers, we get:

t = (80 - 75) / (sqrt((52/25) + (72/25))) = 2.02

Next, a t-table is used to find the critical t-value for the desired level of significance and degrees of freedom (df = n1 + n2 - 2).

Let us assume a significance level of 0.05 and df = 48. The critical t-value is 2.01.

Since the calculated t-value of 2.02 is greater than the critical t-value of 2.01, we can conclude that there is a significant difference between the test scores of students who receive in-person instruction versus those who receive online instruction.

**Example 2:** A researcher wants to know if there is a significant difference in the weight of newborn babies between two hospitals in a city. The researcher randomly selects 20 newborns from Hospital A and 20 newborns from Hospital B and records their weights in pounds. The mean weight for the Hospital A group is 7.5, with a standard deviation of 0.8. The mean weight for the Hospital B group is 7.1, with a standard deviation of 1.2. Is there a significant difference between the two hospitals?

**Solution:** This is an independent samples t-test problem since the two groups being compared are independent of each other. To perform the t-test, we follow the same steps as in Example 1.

Using the formula for the t-value, we get:

t=(7.5-7.1)/(sqrt((0.82/20)+(1.22/20)))  
t = 1.77

Assuming a significance level of 0.05 and df = 38, the critical t-value is 2.024.

Since the calculated t-value of 1.77 is less than the critical t-value of 2.024, we can conclude that there is not a significant difference in the weight of newborn babies between the two hospitals in the city.

## Uses of t-tests

T-test Formula is a widely used statistical tool that helps researchers in many fields to make sense of their data.

* It is used to determine whether two sets of data are significantly different from each other.
* It is used to evaluate if the means of the two groups of data are statistically dissimilar from each other.
* T-tests are used to compare the performance of students in different classes or schools.
* T-tests are used to compare the effectiveness of different marketing strategies.

Overall, t-test formula is a valuable tool for researchers in many fields who want to compare two groups of data and determine if there is a statistically significant difference between them.

## Calculation of t-test

For calculating a t-test, we require three key data values, which are

* The average values from each data set, known as the mean difference.
* The standard deviation of the group.
* The number of key data values of each group.

The result value of the t-test formula gives the t-value. This value is compared against the value of the t-distribution table value. The t-test helps us determine whether the difference is a true difference or an arbitrary, negligible difference.

## Necessary conditions for the application of t-test

The conditions that are necessary to apply T-test are as follows –

* The sample size should be small.
* Statistic follows a normal distribution.
* The value of scaling terms is known.
* Comparison is only between two groups.

## Things to remember

* A t-test is used to verify if there is a considerable difference between the averages of two groups.
* There are two types of t-tests: the independent samples t test formula and the paired samples t-test.
* The t-test is a test used when the objective is hypothesis testing in statistics.
* For an independent samples t-test formula: t = (mean1 - mean2) / (sqrt((s12/n1) + (s22/n2)))
* A greater t-value represents a significant difference between the two groups.
* A smaller t-value represents the difference between the two groups is negligible.

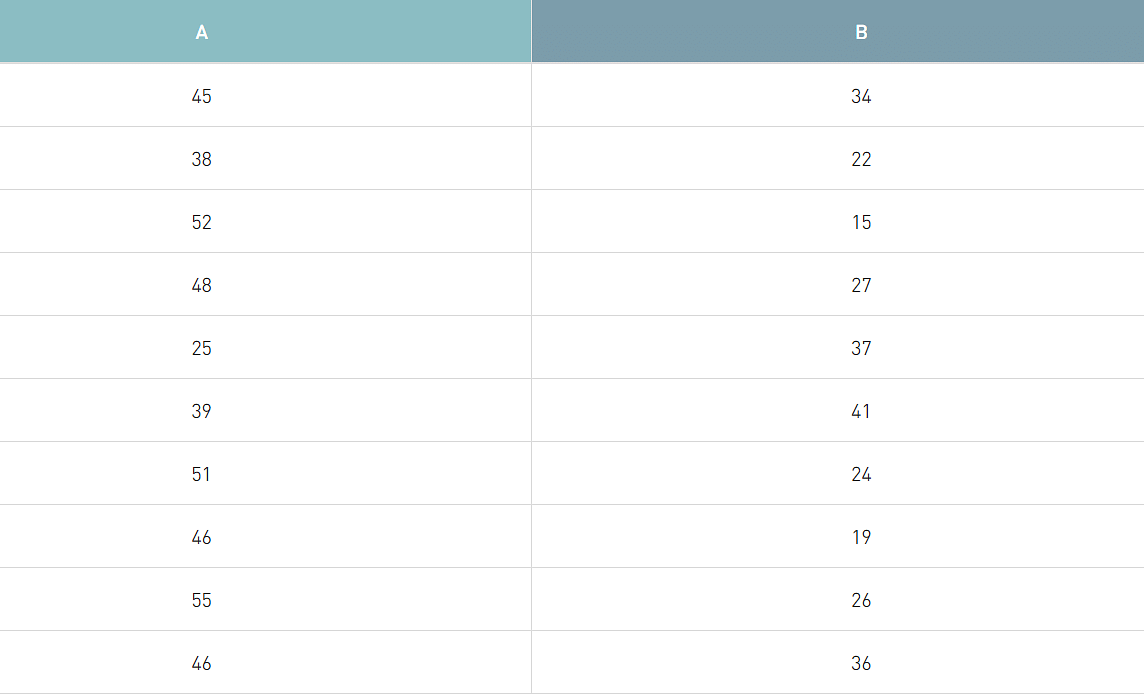
## Sample Questions

**Ques. A professor wants to know if her introductory statistics class has a good grasp of basic math. Six students are chosen at random from the class and given a math proficiency test. The professor wants the class to be able to score above 70 on the test. The six students get scores of 62, 92, 75, 68, 83, and 95. Can the professor have 90 percent confidence that the mean score for the class on the test would be above 70?**

**Ques. A Little League baseball coach wants to know if his team is representative of other teams in scoring runs. Nationally, the average number of runs scored by a Little League team in a game is 5.7. He chooses five games at random in which his team scored 5, 9, 4, 11, and 8 runs. Is it likely that his team's scores could have come from the national distribution? Assume an alpha level of 0.05.**

**Ques. Calculate the t-test for the following data.  
**

**Ques. There are two sets: (2,7,8,9) and (4,3,2,1). Calculate t-test for the same.**

**Ques. Calculate t-test for the following data.   
**

## **Introduction of F Test Formula**

The F Test Formula is a Statistical Formula used to test the significance of differences between two groups of Data. It is often used in research studies to determine whether the difference in the means of two populations is Statistically significant. It is based on the F Statistic, which is a measure of how much variation exists in one group of Data compared to another.

### **What is the Definition of F-Test Statistic Formula?**

### It is a known fact that Statistics is a branch of Mathematics that deals with the collection, classification and representation of Data. F Test is usually used as a generalized Statement for comparing two variances. F Test Statistic Formula is used in various other tests such as regression analysis, the chow test and Scheffe test. F Tests can be conducted by using several technological aid.

### **Definition of F-Test Formula**

F Test is a test Statistic that has an F distribution under the null hypothesis. It is used in comparing the Statistical model with respect to the available Data set. The name for the test is given in honor of Sir. Ronald A Fisher by George W Snedecor. To perform an F Test using technology, the following aspects are to be taken care of.

* State the null hypothesis along with the alternative hypothesis.
* Compute the value of ‘F’ with the help of the standard Formula.
* Determine the value of the F Statistic. The ratio of the variance of the group of means to the mean of the within-group variances.
* As the last step, support or reject the Null hypothesis.

### **F-Test Equation to Compare Two Variances:**

In Statistics, the F-test Formula is used to compare two variances, say σ1 and σ2, by dividing them. As the variances are always positive, the result will also always be positive. Hence, the F -Test equation used to compare two variances is given as:

F\_value =variance1/variance2

s. However, the manual calculation is a little complex and time-consuming.

i.e. F\_value =

σ12/σ22

F Test Formula helps us to compare the variances of two different sets of values. To use F distribution under the null hypothesis, it is important to determine the mean of the two given observations at first and then calculate the variance.

σ2=∑(x−x¯)2/n−1

In the above formula,

σ2 is the variance

x is the values given in a set of data

x is the mean of the given Data set

n is the total number of values in the Data set

While running an F Test, it is very important to note that the population variances are equal. In more simple words, it is always assumed that the variances are equal to unity or 1. Therefore, the variances are always equal in the case of the null hypothesis.

### F Test Statistic Formula Assumptions

F Test equation involves several assumptions. In order to use the F - test Formula, the population should be distributed normally. The samples considered for the test should be independent events. In addition to these, it is also important to consider the following points.

* Calculation of right-tailed tests is easier. To force the test into a right-tailed test, the larger variance is pushed in the numerator.
* In the case of two-tailed tests, alpha is divided by two prior to the determination of critical value.
* Variances are the squares of the standard deviations.

If the obtained degree of freedom is not listed in the F table, it is always better to use a larger critical value to decrease the probability of type 1 errors.

### F-Value Definition: Example Problems

**Example 1:**

Perform an F Test for the following samples.

1. Sample 1 with variance equal to 109.63 and sample size equal to 41.
2. Sample 2 with variance equal to 65.99 and sample size equal to 21.

**Solution:**

**Step 1:**

The hypothesis Statements are written as:

H\_0: No difference in variances

H\_a: Difference invariances

**Step 2:**

Calculate the value of F critical. In this case, the highest variance is taken as the numerator and the lowest variance in the denominator.

F\_value =

σ21 /σ22

F\_value =

109.63/65.99

F\_value = 1.66

**Step 3:**

The next step is the calculation of degrees of freedom.

The degrees of freedom is calculated as Sample size - 1

The degree of freedom for sample 1 is 41 -1 = 40.

The degree of freedom for sample 2 is 21 - 1 = 20.

**Step 4:**

There is no alpha level described in the question, and hence a standard alpha level of 0.05 is chosen. During the test, the alpha level should be reduced to half the initial value, and hence it becomes 0.025.

**Step 5:**

Using the F table, the critical F value is determined with alpha at 0.025. The critical value for (40, 20) at alpha equal to 0.025 is 2.287.

**Step 6:**

It is now the time for comparing the calculated value with the standard value in the table. Generally, the null hypothesis is rejected if the calculated value is greater than the table value. In this F value definition example, the calculated value is 1.66, and the table value is 2.287.

It is clear from the values that 1.66 < 2.287. Hence, the null hypothesis cannot be rejected.

#### Example #1

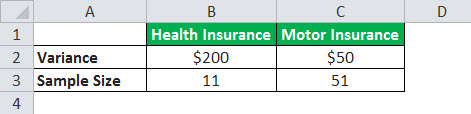
**A statistician was carrying out an F-Test. He got the F statistic as 2.38. The degrees of freedom obtained by him were 8 and 3. Find out the F value from the F Table and determine whether we can reject the null hypothesis at a 5% significance level (one-tailed test).**

Solution:

We must look for 8 and 3 degrees of freedom in the F Table. The F critical value obtained from the table is **8.845**. Since the F statistic (2.38) is lesser than the F Table Value (8.845), we cannot reject the null hypothesis.

#### Example #2

**An insurance company sells health insurance and motor insurance policies. Customers pay premiums for these policies. The CEO of the insurance company wonders if premiums paid by either of the insurance segments (health insurance and motor insurance) are more variable than another. He finds the following data for premiums paid:**



**Conduct a two-tailed F-test with a level of significance of 10%.**

Solution:

* **Step 1:** Null Hypothesis H0: σ12= σ22

Alternate Hypothesis Ha: σ12≠ σ22

* **Step 2:** F statistic = F Value = σ12 / σ22= 200/50 = **4**
* **Step 3:**df1 = n1– 1 = 11-1 =10

df2 = n2– 1 = 51-1 = 50

* **Step 4:** Since it is a two-tailed test, alpha level = 0.10/2 = 0.050. The F value from the F Table with degrees of freedom as 10 and 50 is 2.026.
* **Step 5:**Since F statistic (4) is more than the table value obtained (2.026), we reject the null hypothesis.

#### Example #3

**The bank has a head office in Delhi and a branch in Mumbai. There are long customer queues at one office, while customer queues are short at the other. The Operations Manager of the bank wonders if the customers at one branch are more variable than the number of customers at another. He carries out a research study of customers.**

**The variance of Delhi head office customers is 31, and that for the Mumbai branch is 20. The sample size for the Delhi head office is 11, and that for the Mumbai branch is 21. Carry out a two-tailed F-test with a level of significance of 10%.**

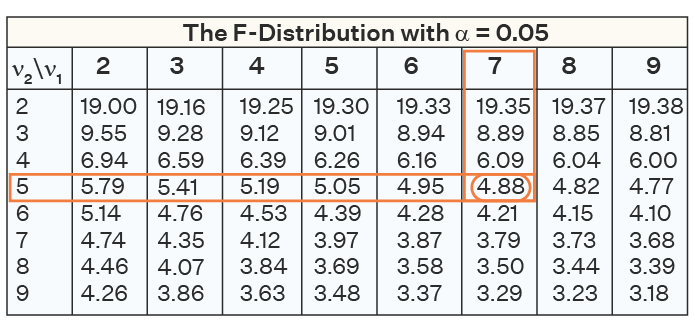
Solution:

* **Step 1:** Null Hypothesis H0: σ12= σ22

Alternate Hypothesis Ha: σ12≠ σ22

* **Step 2:** F statistic = F Value = σ12 / σ22= 31/20 = **1.55**
* **Step 3:**df1 = n1– 1 = 11-1 = 10

df2 = n2– 1 = 21-1 = 20

* **Step 4:**Since it is a two-tailed test, alpha level = 0.10/2 = 0.05. The F value from the F Table with degrees of freedom as 10 and 20 is 2.348.
* **Step 5:** Since the F statistic (1.55) is lesser than the table value obtained (2.348), we cannot reject the null hypothesis.
* **Example 1:** A research team wants to study the effects of a new drug on insomnia. 8 tests were conducted with a variance of 600 initially. After 7 months 6 tests were conducted with a variance of 400. At a significance level of 0.05 was there any improvement in the results after 7 months?  
  **Solution:**As the variance needs to be compared, the f test needs to be used.  
  H0 : s21=s22  
  H1 : s21>s22  
  n1 = 8, n2 = 6  
  df1 = 8 - 1 = 7  
  df2 = 6 - 1 = 5  
  s21 = 600, s22= 400  
  The f test formula is given as follows:  
  F = s21/s22 = 600 / 400  
  F = 1.5  
  Now from the F table the critical value F(0.05, 7, 5) = 4.88  
    
  As 1.5 < 4.88, thus, the null hypothesis cannot be rejected and there is not enough evidence to conclude that there was an improvement in insomnia after using the new drug.  
  **Answer:** Fail to reject the null hypothesis.
* **Example 3:** A toy manufacturer wants to get batteries for toys. A team collected 41 samples from supplier A and the variance was 110 hours. The team also collected 21 samples from supplier B with a variance of 65 hours. At a 0.05 alpha level determine if there is a difference in the variances.  
  **Solution:** This is an example of a two-tailed F test. Thus, the alpha level is 0.05 / 2 = 0.025  
  H0: s21=s22  
  H1: s21≠s22  
  n1= 41, n2 = 21  
  df1 = 41 - 1 = 40  
  df2 = 21 - 1 = 20  
  s21 = 110, s22 = 65  
  F = s21/s22 = 110 / 65  
  F = 1.69  
  Using the F table F(0.025, 40, 20) = 2.287  
  As 1.69 < 2.287 thus, the null hypothesis cannot be rejected,  
  **Answer:** Fail to reject the null hypothesis.

## What is F Test in Statistics?

F test is statistics is a test that is performed on an f distribution. A two-tailed f test is used to check whether the variances of the two given samples (or populations) are equal or not. However, if an f test checks whether one population variance is either greater than or lesser than the other, it becomes a one-tailed hypothesis f test.

### F Test Definition

F test can be defined as a test that uses the f test statistic to check whether the variances of two samples (or populations) are equal to the same value. To conduct an f test, the population should follow an f distribution and the samples must be independent events. On conducting the hypothesis test, if the results of the f test are statistically significant then the null hypothesis can be rejected otherwise it cannot be rejected.

## F Test Formula

The f test is used to check the equality of variances using [hypothesis testing](https://www.cuemath.com/data/hypothesis-testing/). The f test formula for different hypothesis tests is given as follows:

**Left Tailed Test:**

Null Hypothesis: H0 : σ21=σ22

Alternate Hypothesis: H1: σ21<σ22

Decision Criteria: If the f statistic < f critical value then reject the null hypothesis

**Right Tailed test:**

Null Hypothesis: H0: σ21=σ22

Alternate Hypothesis: H1: σ21>σ22

Decision Criteria: If the f test statistic > f test critical value then reject the null hypothesis

**Two Tailed test:**

Null Hypothesis: H0: σ21=σ22

Alternate Hypothesis: H1: σ21≠σ22

Decision Criteria: If the f test statistic > f test critical value then the null hypothesis is rejected

### F Statistic

The f test statistic or simply the f statistic is a value that is compared with the critical value to check if the null hypothesis should be rejected or not. The f test statistic formula is given below:

F statistic for large samples: F = σ21/σ22, where σ21 is the variance of the first population and σ22 is the variance of the second population.

F statistic for small samples: F = s21/s22, where s21 is the variance of the first sample and s22 is the variance of the second sample.

The selection criteria for the σ21 and σ22 for an f statistic is given below:

* For a right-tailed and a two-tailed f test, the variance with the greater value will be in the numerator. Thus, the sample corresponding to σ21will become the first sample. The smaller value variance will be the denominator and belongs to the second sample.
* For a left-tailed test, the smallest variance becomes the numerator (sample 1) and the highest variance goes in the denominator (sample 2).

## F Test Critical Value

A critical value is a point that a test statistic is compared to in order to decide whether to reject or not to reject the null hypothesis. Graphically, the critical value divides a distribution into the acceptance and rejection regions. If the test statistic falls in the rejection region then the null hypothesis can be rejected otherwise it cannot be rejected. The steps to find the f test critical value at a specific alpha level (or significance level), α, are as follows:

* Find the degrees of freedom of the first sample. This is done by subtracting 1 from the first sample size. Thus, x = n1−1.
* Determine the degrees of freedom of the second sample by subtracting 1 from the sample size. This given y = n2−1.
* If it is a right-tailed test then α is the significance level. For a left-tailed test 1 - α is the alpha level. However, if it is a two-tailed test then the significance level is given by α / 2.
* The F table is used to find the critical value at the required alpha level.
* The intersection of the x column and the y row in the f table will give the f test critical value.