How to calculate the P-Value of an F-Statistic in Excel?

Authored by

stats writer

December 27, 2025

The calculation of the P-value associated with an F-statistic is a fundamental requirement in various statistical analyses, particularly in the context of linear regression model testing and Analysis of Variance (ANOVA). Microsoft Excel provides robust built-in functions to facilitate this calculation efficiently. Specifically, the function utilized for this purpose is related to the F-distribution, often denoted as the 'F.DIST' family of functions. This powerful tool takes three crucial inputs: the calculated F-statistic value itself, the numerator degrees of freedom, and the denominator degrees of freedom. The resulting output is the exact probability, or P-value, that corresponds to the observed F-value under the null hypothesis. Understanding how to execute this calculation and interpret the resulting P-value is essential for determining whether the observed F-statistic is considered statistically significant, thereby validating the overall model fit or the comparison between variances.

Introduction to F-Statistics and P-Values

The F-test is a statistical test based on the F-distribution, primarily used to compare variances or assess the overall significance of a regression model containing multiple predictors. When conducting an F-test, the resulting figure is the F-statistic, which quantifies the ratio of variance explained by the model to the unexplained variance (error). This statistic is merely a point estimate; to determine its relevance in the context of hypothesis testing, we must calculate the associated P-value. The P-value serves as the probability of observing a result as extreme as, or more extreme than, the result observed, assuming the null hypothesis is true. A small P-value indicates strong evidence against the null hypothesis, suggesting that the model or variance difference is statistically meaningful.

Calculating this probability accurately is non-trivial without computational aid, as it requires referencing the F-distribution curve which changes shape based on the two parameters known as degrees of freedom. Fortunately, Excel simplifies this process immensely. For virtually all applications of the F-test in regression and ANOVA, we are interested in the right-tailed probability, meaning the area under the F-distribution curve that is greater than or equal to the calculated F-statistic. This area represents the probability of error if we reject the null hypothesis.

The Role of the F-Distribution in Statistical Testing

The F-distribution, named after Sir Ronald Fisher, is a continuous probability distribution that arises frequently as the null distribution of a test statistic in procedures like ANOVA and regression analysis. Its fundamental characteristic is that it is defined by two distinct types of degrees of freedom: the numerator degrees of freedom (df1) and the denominator degrees of freedom (df2). These degrees of freedom dictate the precise shape of the distribution, which is always non-negative and asymmetric. In a linear regression model context, the numerator degrees of freedom

correspond to the number of predictor variables (k) used in the model, while the denominator degrees of freedom relate to the residual error (N - k - 1, where N is the total number of observations).

The position of the calculated F-statistic along this specific distribution curve determines how extreme that value is. If the F-statistic falls far out into the right tail of the distribution, it suggests a low probability of observing such a high ratio of explained variance to unexplained variance purely by chance, assuming the null hypothesis is true. Consequently, a large F-statistic usually translates into a small P-value. When the F-statistic is closer to 1, the explained and unexplained variances are roughly equal, suggesting the model is not much better than random chance, resulting in a large P-value and a failure to reject the null hypothesis.

Calculating the P-Value using Excel's F.DIST.RT Function

To efficiently determine the right-tailed probability associated with an F-statistic in Excel--which is the necessary calculation for standard hypothesis testing--the function F.DIST.RT is utilized. This function is specifically designed to calculate the upper-tail probability (the right-tail area) of the F-distribution. If one were to use the standard F.DIST function, an additional calculation (1 - F.DIST) would be necessary to obtain the correct P-value, but F.DIST.RT simplifies this process by providing the exact P-value directly.

The structure of the command requires precise input of the three required arguments. These arguments must be entered in the correct order to ensure accurate computation of the tail probability. The generic syntax for finding the P-value associated with an F-statistic in Excel is presented below, detailing the variable placeholders that must be replaced by the specific values derived from your statistical analysis:

=F.DIST.RT(x, degree_freedom1, degree_freedom2)

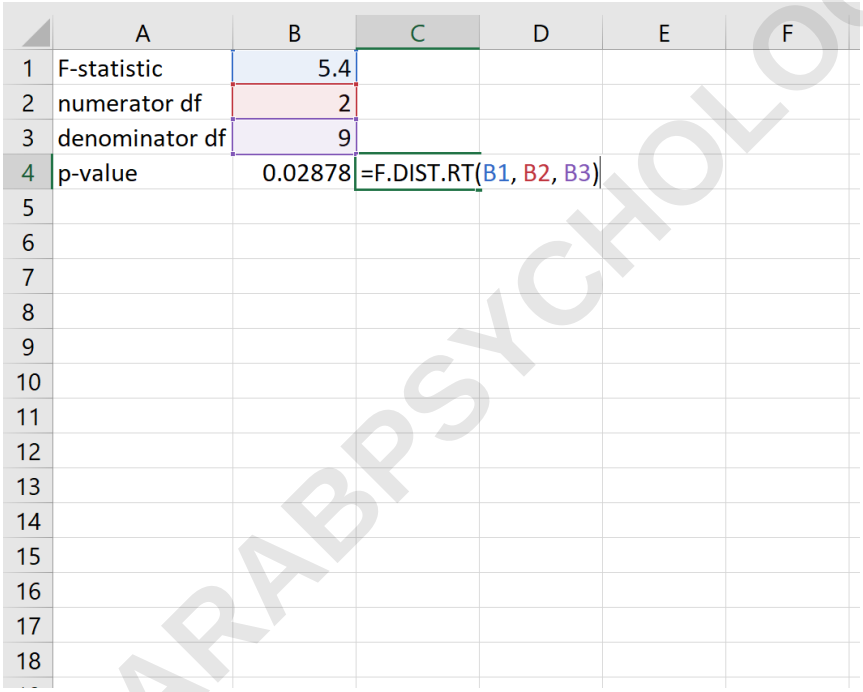
where the variables correspond to the following statistical measurements:

x: This input represents the exact numerical value of the calculated F-statistic derived from the analysis, such as an ANOVA test or a linear regression model output.
degree_freedom1: This is the numerical value for the numerator degrees of freedom (df1), which typically corresponds to the degrees of freedom associated with the model or between-group variance.
degree_freedom2: This is the numerical value for the denominator degrees of freedom (df2), often referred to as the residual or error degrees of freedom, associated with the within-group variance or the total error variance.

Practical Application: A Simple F-Test Example

To illustrate the direct application of the F.DIST.RT function, consider a scenario where an initial statistical test, perhaps an equality of variances test, yields a specific F-statistic. Suppose we obtain an F-statistic value of 5.4. Furthermore, our analysis indicates that the numerator degrees of freedom (df1) is 2, and the denominator degrees of freedom (df2) is 9. We are interested in determining the exact probability associated with observing an F-statistic of 5.4 or higher under the null hypothesis.

In this specific example, the formula entered into any cell in Excel would be as follows: `=F.DIST.RT(5.4, 2, 9)`. This calculation instructs Excel to look up the probability density function for an F-distribution defined by 2 and 9 degrees of freedom, and then calculate the area under the curve to the right of the value 5.4. This is a crucial step for correctly interpreting the outcome of the statistical test.



Executing this function reveals the resulting P-value is approximately **0.02878**. Since this value is below the conventional significance level of 0.05, we would conclude that the observed F-statistic is statistically significant, leading to the rejection of the null hypothesis in favor of the alternative. This basic demonstration confirms the ease with which Excel can be used to handle complex statistical distribution calculations, allowing researchers to focus on interpretation rather than manual probability calculation.

F-Statistics in Multiple Linear Regression Analysis

One of the most frequent and powerful applications of the F-test is in evaluating the overall fit of a multiple linear regression model. When performing multiple regression, the F-statistic tests the null hypothesis that all regression coefficients (excluding the intercept) are simultaneously equal to zero. In simpler terms, it tests whether the entire model, collectively, has any predictive power whatsoever. A significant F-test implies that at least one of the independent variables contributes significantly to explaining the variance in the dependent variable.

The F-statistic in regression is calculated as the ratio of the Mean Square Regression (MSR) to the Mean Square Error (MSE). The MSR represents the variation explained by the model, normalized by the number of predictors, while the MSE represents the unexplained variation, normalized by the error degrees of freedom. A value significantly larger than 1 suggests that the model explains substantially more variance than what is left unexplained by random error. Understanding this context is crucial when attempting to calculate the P-value of the F-statistic for a regression model, which is the focus of the following detailed example.

Case Study: Analyzing a Multivariate Regression Model

To provide a comprehensive illustration, let us analyze a dataset where the goal is to predict a student's final exam score based on two independent variables: the total number of hours studied and the total number of preparation exams taken. This scenario involves 12 different students, providing us with a small, manageable dataset for a multiple regression model. The response variable is the final exam score, while the explanatory variables are study hours and prep exams. This setup allows us to assess the combined explanatory power of both predictors.

The initial dataset, organized within an Excel spreadsheet, forms the foundation of our analysis. It is imperative that the data is correctly structured, with each row representing an observation (student) and separate columns dedicated to the dependent variable (score) and the two independent variables (study hours and prep exams). The structure of this raw data is visually represented in the image below, confirming the inputs that will feed into Excel's powerful Data Analysis ToolPak feature for regression computation.

	A	B	C	D
1	**study_hours**	**prep_exams**	**score**	
2	3	2	76	
3	7	6	88	
4	16	5	96	
5	14	2	90	
6	12	7	98	
7	7	4	80	
8	4	4	86	
9	19	2	89	
10	4	8	68	
11	8	4	75	
12	8	1	72	
13	3	3	76	
14				
15				
16				
17				

Deciphering Excel's Automated Regression Summary Output

When we fit a multiple linear regression model to this data--using **study_hours** and **prep_exams** as the predictor variables and **score** as the response variable--Excel generates a detailed summary table. This table includes several key statistical measures, including R-squared, coefficients, and, most importantly for our purpose, the ANOVA table which contains the F-statistic. The automated calculation is extremely convenient, providing the F-statistic and its associated P-value instantly.

Upon reviewing the ANOVA section of the regression output summary, we extract the critical components necessary for interpreting the overall model fit. The generated output reveals that the F-statistic for the overall regression model is calculated to be **5.0905**. Furthermore, the table provides the corresponding degrees of freedom: 2 for the numerator (corresponding to the two predictor variables) and 9 for the denominator (the residual degrees of freedom, calculated as N-k-1 = 12-2-1 = 9).

	A	B	C	D	E	F	G	H	I	J	K
1	**study_hours**	**prep_exams**	**score**								
2	3	2	76		SUMMARY OUTPUT						
3	7	6	88								
4	16	5	96		*Regression Statistics*						
5	14	2	90		Multiple R	0.7286					
6	12	7	98		R Square	0.5308					
7	7	4	80		Adjusted R Square	0.4265					
8	4	4	86		Standard Error	7.3268					
9	19	2	89		Observations	12					
10	4	8	68								
11	8	4	75		ANOVA						
12	8	1	72			*df*	*SS*	*MS*	*F*	*Significance F*	
13	3	3	76		Regression	2	546.5331	273.2665	5.0905	0.0332	
14					Residual	9	483.1336	53.6815			
15					Total	11	1029.6667				
16											
17						*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
18					Intercept	66.9901	6.2114	10.7849	0.0000	52.9388	81.0414
19					study_hours	1.2999	0.4170	3.1172	0.0124	0.3566	2.2432
20					prep_exams	1.1173	1.0251	1.0899	0.3041	-1.2018	3.4363
21											
22											

Crucially, Excel automatically performs the P-value calculation using the F-statistic and the degrees of freedom. For the F-statistic of 5.0905, the output immediately shows that the corresponding P-value (labeled as "Significance F") is **0.0332**. This value represents the probability of observing an F-statistic of 5.0905 or higher if there were truly no relationship between the predictors and the response variable.

	A	B	C	D	E	F	G	H	I	J	K
1	**study_hours**	**prep_exams**	**score**								
2	3	2	76		SUMMARY OUTPUT						
3	7	6	88								
4	16	5	96		*Regression Statistics*						
5	14	2	90		Multiple R	0.7286					
6	12	7	98		R Square	0.5308					
7	7	4	80		Adjusted R Square	0.4265					
8	4	4	86		Standard Error	7.3268					
9	19	2	89		Observations	12					
10	4	8	68								
11	8	4	75		ANOVA						
12	8	1	72			*df*	*SS*	*MS*	*F*	*Significance F*	
13	3	3	76		Regression	2	546.5331	273.2665	5.0905	0.0332	
14					Residual	9	483.1336	53.6815			
15					Total	11	1029.6667				
16											
17						*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
18					Intercept	66.9901	6.2114	10.7849	0.0000	52.9388	81.0414
19					study_hours	1.2999	0.4170	3.1172	0.0124	0.3566	2.2432
20					prep_exams	1.1173	1.0251	1.0899	0.3041	-1.2018	3.4363
21											
22											

Manual Verification of the Regression P-Value

Although Excel's regression output provides the P-value automatically, it is highly instructive for verification and deeper understanding to calculate this value manually using the F.DIST.RT function, leveraging the F-statistic and degrees of freedom identified in the ANOVA table. This exercise confirms that the automated output is merely an application of the same formula we are learning to use independently.

We take the calculated F-statistic (x = 5.0905), the numerator degrees of freedom (df1 = 2), and the denominator degrees of freedom (df2 = 9). Plugging these values into the F.DIST.RT function yields the following Excel formula: `=F.DIST.RT(5.0905, 2, 9)`. This calculation should yield a result identical to the "Significance F" value provided in the summary output.

	A	B	C	D	E	F	G	H	I	J	K
1	**study_hours**	**prep_exams**	**score**								
2	3	2	76		SUMMARY OUTPUT						
3	7	6	88								
4	16	5	96		*Regression Statistics*						
5	14	2	90		Multiple R	0.7286					
6	12	7	98		R Square	0.5308					
7	7	4	80		Adjusted R Square	0.4265					
8	4	4	86		Standard Error	7.3268					
9	19	2	89		Observations	12					
10	4	8	68								
11	8	4	75		ANOVA						
12	8	1	72			*df*	*SS*	*MS*	*F*	*Significance F*	
13	3	3	76		Regression	2	546.5331	273.2665	5.0905	0.0332	
14					Residual	9	483.1336	53.6815			
15	**0.0332**	=F.DIST.RT(I13, F13, F14)			Total	11	1029.6667				
16											
17						*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
18					Intercept	66.9901	6.2114	10.7849	0.0000	52.9388	81.0414
19					study_hours	1.2999	0.4170	3.1172	0.0124	0.3566	2.2432
20					prep_exams	1.1173	1.0251	1.0899	0.3041	-1.2018	3.4363
21											
22											

As expected, performing this manual calculation confirms the P-value to be 0.0332. This verification step is valuable for ensuring the integrity of the statistical process and building confidence in utilizing Excel's built-in functions. The fact that the manual calculation matches the linear regression output underscores the consistency and reliability of Excel's statistical computation tools.

Interpreting Statistical Significance

The final and most critical step in this process is the interpretation of the resulting P-value. In our regression example, the P-value of 0.0332 must be compared against a pre-determined significance level, commonly denoted as alpha ($alpha$). If the calculated P-value is less than the significance level (e.g., $alpha = 0.05$), we reject the null hypothesis. The null hypothesis in the overall F-test for regression states that all regression coefficients are zero, meaning the model has no explanatory power.

Since 0.0332 is less than 0.05, we successfully reject the null hypothesis. This means we have sufficient evidence to conclude that the overall linear regression model is statistically significant. That is, the combined effect of study hours and prep exams does significantly predict the final exam score. If the P-value had been greater than 0.05 (e.g., 0.15), we would fail to reject the null hypothesis, concluding that the model is ineffective at explaining the variance in the scores. This robust methodology ensures that conclusions drawn from the data are based on statistical evidence rather than mere correlation.