

How to Easily Perform a Partial F-Test in Excel

Authored by
stats writer

December 5, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Easily Perform a Partial F-Test in Excel*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=105995>

Performing a Partial F-Test in Microsoft Excel is a critical procedure for statisticians and analysts seeking to determine if a subset of predictor variables significantly contributes to the overall fit of a regression model. While Excel's Data Analysis Toolpak offers various regression and F-tests, the partial F-test requires a specific sequence of manual calculations after obtaining the initial regression outputs for both the full and reduced models.

This technique allows users to compare two competing models--one comprehensive (full) and one simplified (reduced)--to statistically assess whether the removal of specific predictors leads to a significant loss of explanatory power. The result of this test provides a clear indication of which model structure is most appropriate for the data, ultimately guiding model selection and interpretation. Mastery of this process ensures that unnecessary or non-significant variables are efficiently pruned from the final analysis, leading to a more parsimonious and robust model.

Understanding the Partial F-Test and Nested Models

A Partial F-Test is specifically employed in multiple linear regression to evaluate whether the coefficients associated with a group of independent variables are simultaneously equal to zero. This procedure is fundamental for determining the statistical significance of adding or removing a block of predictors. The test's utility lies in its ability to compare a "full" model against a "nested" or "reduced" version of that same model, ensuring that the model comparison is valid and interpretable.

The concept of a nested model is central to this test. A model is considered nested within another if it contains a strict subset of the predictor variables present in the larger, comprehensive model. The reduced model must be derived simply by setting one or more coefficients in the full model to zero. If this hierarchical relationship does not exist--meaning the models use entirely different sets of variables or have different fundamental structures--then a partial F-test cannot be legitimately used for comparison.

For illustrative purposes, consider a full model incorporating four predictor variables (x_1 , x_2 , x_3 , x_4). The equation is represented as:

$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \epsilon$$

An appropriate nested model, which removes predictors x_3 and x_4 , would be defined as:

$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \epsilon$$

The Partial F-Test assesses whether the variance explained by the removed terms (x_3 and x_4) is large enough to justify their inclusion in the model. This is equivalent to testing the null hypothesis that $\beta_3 = \beta_4 = 0$.

The Mathematical Basis of the Partial F-Statistic

To determine if these two models (full and reduced) exhibit a statistically significant difference in their explanatory power, we calculate the Partial F test statistic. This statistic is essentially a ratio comparing the increase in the Residual Sum of Squares (RSS) due to model reduction, relative to the unexplained variance (error) remaining in the full model.

The formal equation for calculating the Partial F-statistic is derived by partitioning the total variance explained and adjusting for the respective degrees of freedom:

$$F = ((RSS_{\text{reduced}} - RSS_{\text{full}})/p) / (RSS_{\text{full}}/(n-k))$$

Understanding each component of this formula is vital for accurate computation and interpretation:

RSS_{reduced}: This represents the Residual Sum of Squares obtained from the simpler, nested model. Since it has fewer predictors, this value will always be greater than or equal to RSS_{full}.

RSS_{full}: This is the Residual Sum of Squares from the complete regression model, representing the minimum unexplained variance achievable by the full set of predictors.

p: This denotes the number of predictor variables that were removed from the full model to create the reduced model. This also serves as the numerator degrees of freedom.

n: This is the total number of observations (sample size) included in the dataset used for both regression analyses.

k: This signifies the total number of estimated coefficients in the full model, which includes the intercept (β_0) and all predictor coefficients. The term (n-k) represents the denominator degrees of freedom.

The resulting F-statistic is compared against an F-distribution with degrees of freedom (p, n-k) to determine the associated p-value and make a formal statistical decision regarding the model comparison.

Formulating the Hypotheses

The structure of the Partial F-Test is built upon a standard framework of hypothesis testing. We are testing whether the explanatory power lost by moving from the full model to the reduced model is statistically negligible or significant.

The two hypotheses tested are:

H₀ (Null Hypothesis): All coefficients corresponding to the predictor variables removed from the full model are equal to zero (e.g., $\beta_3 = \beta_4 = 0$). This suggests the reduced model is sufficient and the extra predictors do not significantly improve the fit.

H_A (Alternative Hypothesis): At least one of the coefficients removed from the full model is non-

zero (e.g., $\beta_3 \neq 0$ or $\beta_4 \neq 0$). This suggests the full model provides a statistically better fit than the reduced model, and the removed variables are jointly significant.

If the p-value calculated from the F test-statistic falls below the predefined significance level (commonly $\alpha = 0.05$), we possess sufficient evidence to reject the null hypothesis (H_0). Rejecting H_0 implies that the extra predictors, when considered together, contribute meaningfully to the prediction of the dependent variable, and thus the full model is the preferred choice. Conversely, failure to reject H_0 supports the use of the simpler, reduced model.

Example: Partial F-Test Setup in Excel

The following practical example demonstrates how to set up and execute the necessary steps in Excel to perform a Partial F-test. We must first ensure the data is correctly structured and then run separate regression analyses for both the full and reduced models.

Suppose we have gathered a dataset containing a dependent variable (Y) and four independent predictor variables (x_1, x_2, x_3, x_4), as illustrated below:

	A	B	C	D	E	F	G
1	x1	x2	x3	x4	y		
2	1	15	23	19	18		
3	3	14	25	19	20		
4	3	15	24	17	24		
5	4	12	24	17	24		
6	4	10	14	15	23		
7	5	9	9	18	26		
8	7	6	17	15	29		
9	8	7	17	13	30		
10	8	5	4	19	37		
11	9	4	13	8	34		
12	10	4	6	8	38		
13	12	3	8	17	40		
14							
15							
16							
17							
18							
19							
20							
21							
22							

Our objective is to test whether predictors x_3 and x_4 significantly improve the model fit beyond what is already achieved by x_1 and x_2 . This means we are comparing the predictive power of two specific models:

Full Model: $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4$

Reduced Model: $y = \beta_0 + \beta_1x_1 + \beta_2x_2$

The first step in Excel is to utilize the Data Analysis Toolpak (found under the Data tab) to run the Regression function separately for each of these model specifications. Ensure that the correct input ranges are selected for the Y-variable and the respective X-variables (all four for the full model, only x_1 and x_2 for the reduced model).

Extracting Key Components: Residual Sum of Squares (RSS)

After running the regression analysis for both the full and reduced models, Excel generates extensive output tables. The crucial piece of information needed for the partial F-test calculation is the Residual Sum of Squares (RSS) from the ANOVA section of each regression output. This value represents the variation in the dependent variable that is not explained by the model.

Upon performing regression analysis for both the full and reduced specifications, we obtain the following outputs, which clearly display the necessary ANOVA statistics for both models:

G	H	I	J	K	L	M
Full Model	ANOVA					
		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
	Regression	4	561.6047	140.4012	38.82785	7.24E-05
	Residual	7	25.31194	3.615992		
	Total	11	586.9167			
		<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>
	Intercept	14.34009	7.565093	1.89556	0.099856	-3.54851
	x1	2.290687	0.59909	3.82361	0.006512	0.874064
	x2	0.516173	0.567794	0.909086	0.393528	-0.82645
	x3	-0.29052	0.153701	-1.89018	0.100649	-0.65397
	x4	0.006391	0.189974	0.033642	0.974102	-0.44283
	SUMMARY OUTPUT					
	<i>Regression Statistics</i>					
	Multiple R	0.965112				
	R Square	0.931441				
	Adjusted R	0.916206				
Reduced Model	ANOVA					
		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
	Regression	2	546.6782	273.3391	61.13686	5.78E-06
	Residual	9	40.23844	4.470938		
	Total	11	586.9167			
		<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>
	Intercept	17.50073	7.99721	2.188354	0.056401	-0.59022
	x1	1.965388	0.640132	3.070286	0.013348	0.517309
	x2	-0.11969	0.475321	-0.2518	0.806849	-1.19494

From the generated output tables, we must carefully locate and record the RSS values. The RSS is typically listed under the 'Residual' row in the 'SS' (Sum of Squares) column of the ANOVA table. For instance, based on the outputs provided above, we would determine:

RSS_{full}: The Residual SS from the full model output.

RSS_{reduced}: The Residual SS from the reduced model output.

It is also imperative to identify the degrees of freedom and sample size parameters needed for the denominator of the F-statistic: n (total observations), p (number of removed predictors, here p=2), and k (number of coefficients in the full model, here k=5). These values are essential for

calculating the correct critical F-value and the associated p-value.

Calculating the Partial F-Statistic Manually

Once RSS_{full} , $RSS_{reduced}$, p , n , and k are identified, the calculation of the F-statistic can be executed directly within an Excel spreadsheet using the defined formula. This typically involves setting up a dedicated calculation area where these values are referenced.

The structure of the calculation in Excel should explicitly reflect the ratio formula: the numerator measures the improvement in fit per predictor added (or loss per predictor removed), and the denominator measures the residual variance of the full model (Mean Square Error).

Using Excel functions, we calculate the resulting F-test statistic:

	G	H	I	J	K	L	M	N	O	P	Q	R	S
Full Model	ANOVA									F test statistic	2.064	$= (J12-J34)/2/K13$	
		df	SS	MS	F	Significance F							
	Regression	4	561.6047	140.4012	38.82785	7.24E-05							
	Residual	7	25.31194	3.615992									
	Total	11	586.9167										
		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%				
	Intercept	14.34009	7.565093	1.89556	0.099856	-3.54851	32.22869	-3.54851	32.22869				
	x1	2.290687	0.59909	3.82361	0.006512	0.874064	3.70731	0.874064	3.707309921				
	x2	0.516173	0.567794	0.909086	0.393528	-0.82645	1.858791	-0.82645	1.858791347				
	x3	-0.29052	0.153701	-1.89018	0.100649	-0.65397	0.072923	-0.65397	0.072923129				
	x4	0.006391	0.189974	0.033642	0.974102	-0.44283	0.455609	-0.44283	0.45560912				
	SUMMARY OUTPUT												
	Regression Statistics												
	Multiple R	0.965112											
	R Square	0.931441											
	Adjusted R	0.916206											
	Standard E	2.114459											
	Observations	12											
Reduced Model	ANOVA												
		df	SS	MS	F	Significance F							
	Regression	2	546.6782	273.3391	61.13686	5.78E-06							
	Residual	9	40.23844	4.470938									
	Total	11	586.9167										

Following the arithmetic dictated by the components derived from the regression outputs, the resulting F test statistic in this specific example is calculated to be approximately **2.064**. This calculated F-value must then be evaluated against the appropriate F-distribution to ascertain its statistical significance.

Determining the P-Value and Interpreting Results

The final step involves converting the calculated F-statistic (2.064) into a probability, or p-value, using Excel's statistical functions. The F-distribution is characterized by two degrees of freedom: the numerator degrees of freedom (p, the number of predictors removed) and the denominator degrees of freedom (n-k, the residual degrees of freedom of the full model).

In Excel, the function

`F.DIST.RT(F_statistic, df1, df2)`

is used to calculate the right-tailed probability, which corresponds to the p-value for the F-test. This calculation confirms the probability of observing an F-statistic as extreme as 2.064, assuming the null hypothesis (H0) is true.

G	H	I	J	K	L	M	N	O	P	Q	R	S	T
Full Model	ANOVA								F test statistic	2.064			
		df	SS	MS	F	Significance F			p-value	0.197422	=F.DIST.RT(Q10, I35-I13, I13)		
	Regression	4	561.6047	140.4012	38.82785	7.24E-05							
	Residual	7	25.31194	3.615992									
	Total	11	586.9167										
Reduced Model	ANOVA												
		df	SS	MS	F	Significance F							
	Regression	2	546.6782	273.3391	61.13686	5.78E-06							
	Residual	9	40.23844	4.470938									
	Total	11	586.9167										
Full Model	Coefficients and Standard Error					t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%		
	Intercept	14.34009	7.565093	1.89556	0.099856	-3.54851	32.22869	-3.54851	32.22869	-3.54851	32.22869	351	
	x1	2.290687	0.59909	3.82361	0.006512	0.874064	3.70731	0.874064	3.70731	0.874064	3.70731	9921	
	x2	0.516173	0.567794	0.909086	0.393528	-0.82645	1.858791	-0.82645	1.858791	-0.82645	1.858791	1347	
	x3	-0.29052	0.153701	-1.89018	0.100649	-0.65397	0.072923	-0.65397	0.072923	-0.65397	0.072923	129	
	x4	0.006391	0.189974	0.033642	0.974102	-0.44283	0.455609	-0.44283	0.455609	-0.44283	0.455609	12	
Reduced Model	SUMMARY OUTPUT												
Reduced Model	Regression Statistics												

Based on the specified degrees of freedom, the resulting p-value turns out to be approximately **0.1974**.

To make a definitive decision, we compare the calculated p-value to the chosen significance level ($\alpha = 0.05$). Since 0.1974 is greater than 0.05, we fail to reject the null hypothesis (H0). This crucial finding indicates that there is insufficient statistical evidence to conclude that the coefficients for the removed predictor variables, x3 and x4, are jointly non-zero. In practical terms, this means adding x3 and x4 to the regression model does not result in a statistically significant improvement in the model's ability to fit or predict the outcome. Therefore, the simpler, reduced model is preferable.

based on the principle of parsimony.

Conclusion: Guiding Model Selection

The Partial F-Test serves as an invaluable tool for rigorous model selection in multiple regression analysis. By systematically comparing nested models, analysts can ensure that their final model is not only statistically sound but also as simple and efficient as possible. This prevents overfitting, where unnecessary variables dilute the interpretability and generalizability of the results.

When the test suggests that the additional predictors do not significantly enhance the model fit, the best practice is to adopt the reduced model. Conversely, a significant F-statistic requires the retention of the full model to capture the meaningful contributions of all included variables. Mastering the manual calculation of the partial F-test in Excel provides a deep understanding of variance partitioning and hypothesis testing principles, essential skills for any data professional.