

# Understanding and Performing Partial F-Tests in Excel: A Step-by-Step Guide

Authored by  
**Mohammed looti**

November 4, 2025

## RECOMMENDED CITATION

Mohammed looti (2025). *Understanding and Performing Partial F-Tests in Excel: A Step-by-Step Guide*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=10197>

## Introduction: The Necessity of the Partial F-Test in Regression

The [Partial F-test](#) is an indispensable technique utilized in multivariate statistical analysis to rigorously evaluate the collective contribution of a specific set of predictor variables within a [regression model](#). This test is crucial for determining whether incorporating additional complexity, moving from a reduced (simpler) model to a full (more complex) model, yields a statistically significant improvement in the model's explanatory power. Effective model selection requires balancing high predictive accuracy with parsimony, and the Partial F-test provides the objective, data-driven framework needed to make this decision.

The foundation of this test rests on the comparison of two models: a comprehensive, or **full model**, and a **nested model**. A nested model is defined as one that contains a subset of the predictor variables found within the full model. This hierarchical relationship allows us to isolate the effect of the removed variables. Since adding any predictor, even a useless one, will inevitably increase the  $R^2$  value--making the fit appear better--the Partial F-test provides the statistical rigor necessary to confirm whether this observed improvement is greater than random chance.

To illustrate this concept, consider a scenario where we begin with a comprehensive linear [regression model](#) involving four independent variables:

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

If, for theoretical or practical reasons, we hypothesize that  $x_3$  and  $x_4$  are redundant predictors, we construct the corresponding nested model by constraining their coefficients to zero:

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

The [Partial F-test](#) then formally assesses whether the full model is statistically superior to the reduced model by examining the collective significance of  $\beta_3$  and  $\beta_4$ .

## The Core Formula and Hypothesis Structure

The statistical power of the [Partial F-test](#) derives from comparing the unexplained variance in both models. This variance is precisely measured by the [Residual Sum of Squares \(RSS\)](#). The test statistic transforms the difference in RSS between the two models into a ratio of mean squares, allowing us to test the null hypothesis against the F-distribution. The formula for the F test statistic is structured to compare the marginal reduction in error achieved by the added variables against the baseline error of the full model:

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

The numerator calculates the average reduction in RSS per removed variable, while the

denominator represents the Mean Squared Error (MSE) of the best model (the full model). If the removed variables were indeed influential, the difference  $(RSS_{\text{reduced}} - RSS_{\text{full}})$  would be large, leading to a large F-statistic.

Accurate application of the formula requires precise identification of its components, derived directly from the regression output and the data structure:

**RSS<sub>reduced</sub>**: The [Residual Sum of Squares](#) from the restricted model.

**RSS<sub>full</sub>**: The [Residual Sum of Squares](#) from the complex model.

**p**: The number of predictors removed, which dictates the numerator [degrees of freedom](#).

**n**: The total number of observations (sample size).

**k**: The total number of estimated coefficients in the full model (including the intercept  $\beta_0$ ). This term helps define the denominator [degrees of freedom](#) as  $(n-k)$ .

The formal hypotheses guiding the test are stated as follows:

**H<sub>0</sub> (The Null Hypothesis)**: All coefficients removed from the full model are collectively equal to zero (e.g.,  $\beta_3 = \beta_4 = 0$ ). This suggests the predictors are not needed.

**H<sub>A</sub> (The Alternative Hypothesis)**: At least one of the removed coefficients is non-zero (e.g.,  $\beta_3 \neq 0$  or  $\beta_4 \neq 0$ ). This suggests the full model is significantly better.

## Prerequisites and Data Preparation in Excel

To initiate the Partial F-test procedure within the Excel environment, the **Data Analysis ToolPak** add-in must be enabled. This is accomplished via the 'File' menu, selecting 'Options', then 'Add-ins', and finally activating the ToolPak. This add-in provides the necessary Regression utility needed to generate the ANOVA tables from which the RSS values are extracted. Without the ToolPak, the manual calculation of RSS becomes prohibitively complex and prone to error.

Data preparation involves organizing the dependent variable (Y) and all potential independent variables ( $x_1$  through  $x_4$ ) in adjacent columns. For our illustrative example, we work with a dataset of ten observations, allowing us to test the joint significance of  $x_3$  and  $x_4$  when  $x_1$  and  $x_2$  are already included.

	A	B	C	D	E	F	G
1	<b>x1</b>	<b>x2</b>	<b>x3</b>	<b>x4</b>	<b>y</b>		
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							

The procedure mandates running the Regression Analysis tool exactly twice. The first run establishes the **Full Model**, where the Input X Range includes all predictors ( $x_1$  to  $x_4$ ). The second run establishes the **Reduced Model**, where the Input X Range includes only the predictors we intend to keep ( $x_1$  and  $x_2$ ). Ensuring that the input ranges are correctly specified for each model run is the most critical preparatory step, as any error here will invalidate the resulting RSS values.

## Extracting the Residual Sum of Squares (RSS)

The comparison of the **unexplained variance** constitutes the foundation of the Partial F-test. After executing the two separate regression analyses, two distinct output sheets containing ANOVA tables will be generated. The goal is to carefully locate and record the [Residual Sum of Squares \(RSS\)](#) for each model. This value is always found in the row labeled "Residual" under the column titled "Sum of Squares" within the respective ANOVA sections.

In addition to extracting  $RSS_{\text{full}}$  and  $RSS_{\text{reduced}}$ , we must explicitly confirm the parameters that define the [degrees of freedom](#) for the F-distribution:  $n=10$  total observations,  $k=5$  total coefficients in the full model (intercept + 4 predictors), and  $p=2$  predictors removed ( $x_3$  and  $x_4$ ). These values are constant regardless of the RSS results.

The consolidated output below illustrates the location of the necessary statistics after running both

regressions:

G	H	I	J	K	L	M
<b>Full Model</b>	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% Bound</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				
	Standard Error	2.114459				
	Observations	12				
<b>Reduced Model</b>	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% Bound</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 this output, we successfully identify  $RSS_{\text{reduced}} = 102.50$  and  $RSS_{\text{full}} = 56.40$ . The substantial difference between these two values suggests a potential gain, which must now be statistically tested against the F-distribution to determine if the gain is significant given the inherent variability.

### Calculating the F-Statistic and P-Value

With all parameters secured, the next step is the precise calculation of the F test-statistic within a

new section of the Excel worksheet. This calculation is performed according to the formula:  $F = ((102.50 - 56.40) / 2) / (56.40 / (10 - 5))$ .

The detailed breakdown in Excel, which maintains the required [degrees of freedom](#), results in the final F-statistic:

G	H	I	J	K	L	M	N	O	P	Q	R	S
<b>Full Model</b>	ANOVA								<b>F test statistic</b>	2.064	<code>= (J12-J34)/2/K13</code>	
		<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>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>			
	Intercept	14.34009	7.565093	1.89556	0.099856	-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	9921		
	x2	0.516173	0.567794	0.909086	0.393528	-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	129		
	x4	0.006391	0.189974	0.033642	0.974102	-0.44283	0.455609	-0.44283	0.455609	12		
	SUMMARY OUTPUT											
	<i>Regression Statistics</i>											
	Multiple R	0.965112										
	R Square	0.931441										
	Adjusted R	0.916206										
	Standard E	2.114459										
	Observatio	12										
<b>Reduced Model</b>	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									

The calculated F test-statistic for this specific model comparison is **2.064**. To make a formal decision based on this statistic, we must translate it into a corresponding [p-value](#) using Excel's statistical distribution functions.

We employ the `F.DIST.RT` function, which computes the right-tailed probability of the F-distribution, requiring the calculated F-statistic and the two degrees of freedom ( $p=2$  and  $n-k=5$ ). The right-tailed calculation is essential because we are testing if the F-ratio is significantly large, indicating a strong rejection of the [null hypothesis](#).

The formula used in Excel is `=F.DIST.RT(2.064, 2, 5)`, as demonstrated below:



## Additional Resources for Advanced Statistical Testing

Mastering model selection techniques like the Partial F-test requires a strong background in broader statistical concepts. For analysts seeking to enhance their methodological toolkit beyond nested model comparisons, exploring the following advanced topics is highly beneficial:

A rigorous understanding of the assumptions of Linear Regression (linearity, independence, homoscedasticity, and normality of residuals) is crucial, as violations can undermine the validity of the F-test results.

Methods for comparing non-nested models, such as using the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC), which offer powerful alternatives when models do not share a direct subset relationship.

Techniques for diagnosing and mitigating the effects of multicollinearity among predictors, which can severely inflate standard errors and complicate the interpretation of individual coefficients.