

# Learning About the Null Hypothesis in Linear Regression

Authored by  
**Mohammed loot**

November 4, 2025

## RECOMMENDED CITATION

Mohammed loot (2025). *Learning About the Null Hypothesis in Linear Regression*.  
PSYCHOLOGICAL STATISTICS. Retrieved from  
<https://statistics.arabpsychology.com/?p=9815>

[Linear regression](#) is a cornerstone statistical methodology used extensively to model, predict, and quantify the relationship between one or more predictor variables and a single [response variable](#). The primary statistical objective of this powerful technique is to determine the line or hyperplane that best fits the observed data, thereby summarizing the underlying relationship. This model allows analysts to forecast future outcomes and rigorously measure the specific influence of each predictor factor on the outcome.

Crucially, before asserting that any observed pattern represents a genuine relationship within the wider population, we must subject our findings to formal scrutiny using [hypothesis testing](#). This entire inferential process revolves around the formulation of the [null hypothesis](#) ( $H_0$ ), a statement of no effect or no relationship. In the context of regression analysis, the null hypothesis posits that the predictor variables have no statistically significant influence on the response variable.

## Understanding Simple Linear Regression (SLR)

When the analysis is constrained to examining the relationship between exactly one predictor variable and one response variable, we utilize [simple linear regression](#) (SLR). SLR employs the equation of a straight line to estimate how the response changes relative to the predictor. This mathematical framework provides the simplest yet fundamental means of quantifying linearity in statistical modeling.

The theoretical population model for SLR is expressed as:

$$y = \beta_0 + \beta_1 x$$

The parameters within this equation are defined as follows, representing the essential components required to define the line of best fit:

**$y$** : Represents the estimated or predicted value of the response variable for a specific input value of  $x$ .

**$\beta_0$** : Known as the intercept, this parameter reflects the expected average value of the response variable ( $y$ ) when the predictor variable ( $x$ ) is precisely zero.

**$\beta_1$** : This is the slope [coefficient](#)--the most critical parameter. It quantifies the average change in the response variable ( $y$ ) that is expected to occur for every one-unit increase in the predictor variable ( $x$ ).

**$x$** : Denotes the observed value of the specific predictor variable utilized in the model.

## The Foundational Null Hypothesis in SLR

The central goal of statistical inference in SLR is to definitively ascertain whether the observed

relationship (represented by the calculated sample slope,  $b_1$ ) is strong enough to confidently conclude that a relationship truly exists in the population (represented by the theoretical population slope,  $\beta_1$ ). The hypothesis test in SLR is entirely focused on evaluating the magnitude and significance of the population slope coefficient  $\beta_1$ .

The formal hypotheses utilized in simple linear regression analysis are structured as follows:

**H<sub>0</sub> (Null Hypothesis):**  $\beta_1 = 0$

**H<sub>A</sub> (Alternative Hypothesis):**  $\beta_1 \neq 0$

The [null hypothesis](#) ( $H_0$ ) asserts that the population slope coefficient ( $\beta_1$ ) is exactly equal to zero. If the slope is zero, it directly implies that fluctuations in the predictor variable ( $x$ ) have absolutely no systematic impact on the response variable ( $y$ ). Consequently, failing to reject the null hypothesis suggests that there is no statistically significant linear relationship between  $x$  and  $y$ . Conversely, the alternative hypothesis ( $H_A$ ) maintains that the slope ( $\beta_1$ ) is **not** equal to zero. When we reject  $H_0$  in favor of  $H_A$ , we are concluding that the predictor variable contributes meaningfully to explaining the variation in the response variable, confirming the existence of a [statistically significant relationship](#).

## Extending the Concept: Multiple Linear Regression (MLR)

When the predictive model incorporates two or more predictor variables acting simultaneously upon a single response variable, the technique is extended to [multiple linear regression](#) (MLR). This methodology is essential for developing a more complex and realistic understanding of phenomena where multiple independent factors jointly influence the outcome of interest.

The MLR formula generalizes the simple model to accommodate  $k$  number of predictor variables:

$$\hat{y} = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_kx_k$$

The interpretation of the coefficients in MLR is subtly refined compared to SLR, reflecting the complexity introduced by having multiple predictors:

**$\hat{y}$ :** The estimated response value derived from the combined influence of all predictor variables.

**$\beta_0$ :** The average value of  $y$  when all predictor variables ( $x_1$  through  $x_k$ ) are held at zero.

**$\beta_i$ :** Represents the average change in  $y$  associated with a one-unit increase in  $x_i$ , critically assuming that all other predictor variables in the model are held constant (the principle known as **ceteris paribus**).

**$x_i$ :** The observed value of the specific predictor variable  $x_i$ .

MLR necessitates performing a global test to determine the overall utility and significance of the

entire model. This test evaluates whether the entire ensemble of predictors, when considered collectively, significantly explains the variation observed in the response variable. The hypotheses for this crucial overall model significance test are:

**H0 (Global Null Hypothesis):**  $\beta_1 = \beta_2 = \dots = \beta_k = 0$

**HA (Global Alternative Hypothesis):** At least one  $\beta_i \neq 0$

The global null hypothesis is a stringent statement that declares that **every single coefficient** in the model is simultaneously zero. Acceptance of this null hypothesis would lead to the conclusion that the model is effectively worthless; none of the predictor variables, individually or combined, possess a statistically significant ability to explain the variance in the response variable,  $y$ . Conversely, the global alternative hypothesis is satisfied if even one of the regression coefficients is proven to be non-zero, indicating that the MLR model, as a whole, provides a statistically useful level of predictive power.

## Case Study 1: Testing Significance in Simple Linear Regression

Let us examine a practical scenario where a statistics professor aims to predict a student's final exam score based solely on the number of hours they dedicated to studying. Data is collected from a sample of 20 students, and a [simple linear regression](#) model is fitted to the data.

The primary inferential task is to determine if the variable "hours studied" is a statistically significant predictor of "exam score." We formally test the null hypothesis:  $H_0: \beta_1 = 0$ .

The following output summarizes the key findings derived from the regression model analysis:

D	E	F	G	H	I	J	K	L
SUMMARY OUTPUT								
<i>Regression Statistics</i>								
Multiple R	0.8528							
R Square	0.7273							
Adjusted R Square	0.7121							
Standard Error	5.2805							
Observations	20							
ANOVA								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	1	1338.2906	1338.2906	47.9952	0.0000			
Residual	18	501.9094	27.8839					
Total	19	1840.2000						
	<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	67.1617	2.6633	25.2178	0.0000	61.5664	72.7570	61.5664	72.7570
hours	5.2503	0.7578	6.9279	0.0000	3.6581	6.8424	3.6581	6.8424

Based upon the estimated coefficients provided in this output, the specific simple linear regression model derived from the sample data is:

$$\text{Exam Score} = 67.1617 + 5.2503 * (\text{hours studied})$$

To formally evaluate the null hypothesis ( $H_0$ ), we must scrutinize the overall model fit using the [F-Value](#) (sourced from the Analysis of Variance, or ANOVA table) and its corresponding [p-value](#):

Overall F-Value: **47.9952**

P-value: **0.000**

In standard statistical testing, we establish a critical significance level (alpha,  $\alpha$ ) typically set at 0.05. Since the calculated p-value (0.000) is substantially smaller than  $\alpha$  (0.05), we have compelling evidence to reject the null hypothesis ( $H_0: \beta_1 = 0$ ). This decisive rejection permits us to conclude that there is a **statistically significant relationship** between the number of hours studied and the exam score achieved. Furthermore, the model suggests that for every additional hour studied, the predicted exam score increases by an average of 5.2503 points.

## Case Study 2: Interpreting the Multiple Linear Regression F-Test

To enrich the predictive power, let us expand the previous study by incorporating a second predictor: the number of preparatory exams taken. The professor now fits a [multiple linear regression](#) model to the data, utilizing both "hours studied" ( $x_1$ ) and "prep exams taken"

( $x_2$ ) to predict "exam score" ( $y$ ).

In this MLR context, we conduct a global F-test to verify if the two predictors, in combination, offer a statistically significant prediction of the response variable. The global null hypothesis under consideration is  $H_0: \beta_{\text{hours}} = \beta_{\text{prep}} = 0$ .

The regression output detailing the overall model fit is presented below:

D	E	F	G	H	I	J	K
SUMMARY OUTPUT							
<i>Regression Statistics</i>							
Multiple R	0.857						
R Square	0.734						
Adjusted R Square	0.703						
Standard Error	5.366						
Observations	20						
ANOVA							
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>		
Regression	2	1350.76	675.38	23.46	0.00		
Residual	17	489.44	28.79				
Total	19	1840.20					
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	
Intercept	67.67	2.82	24.03	0.00	61.73	73.61	
hours	5.56	0.90	6.18	0.00	3.66	7.45	
prep_exams	-0.60	0.91	-0.66	0.52	-2.53	1.33	

The resulting fitted multiple linear regression model equation based on this output is:

$$\text{Exam Score} = 67.67 + 5.56 * (\text{hours studied}) - 0.60 * (\text{prep exams taken})$$

To rigorously determine if there is a **jointly statistically significant relationship** between these two predictors and the response variable, we must analyze the model's overall F-statistic and its corresponding p-value:

Overall F-Value: **23.46**

P-value: **0.00**

Given that the overall p-value (0.00) is definitively lower than the standard significance level ( $\alpha = 0.05$ ), we must confidently reject the global null hypothesis ( $H_0$ ). This statistical conclusion signifies that the combination of predictors--hours studied and prep exams taken--

possesses a jointly statistically significant relationship with the exam score. Fundamentally, this confirms that the MLR model is useful and provides predictive value.

**Important Note on Interpretation:** It is vital for analysts to differentiate between the conclusion of the overall model test (the F-test) and the tests for individual coefficient significance (t-tests). While the overall model is deemed significant, observing the individual coefficient output reveals that the p-value for the "prep exams taken" coefficient ( $p = 0.52$ ) is **not** individually significant (as  $0.52 > 0.05$ ). However, because the F-test assesses the variables **combined as a group**, the rejection of the global  $H_0$  confirms that the set of predictors, taken together, has a significant predictive relationship with the exam score, even if one component does not contribute uniquely when the influence of the other is held constant.

## Advanced Considerations and Further Study

The null hypothesis remains the critical starting point for all inferential statistical analyses, particularly in linear regression. A thorough grasp of these concepts is essential for moving beyond simply fitting a line to data towards making robust conclusions about population relationships. For those aiming for a deeper understanding of the underlying mathematical principles, including the calculation of the F-statistic and the mechanics of regression coefficients, further exploration is highly recommended. Developing a strong comprehension of related concepts such as [statistical power](#) and the definitions of [Type I and Type II errors](#) will provide superior clarity on why the null hypothesis serves as the cornerstone of reliable regression analysis.