

How to Define the Null Hypothesis in Linear Regression

Authored by
stats writer

December 5, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Define the Null Hypothesis in Linear Regression*.
PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=105578>

The null hypothesis for linear regression is that there is no relationship between the independent variable and the dependent variable, meaning that any observed difference between the two is due to chance.

Linear regression is a technique we can use to understand the relationship between one or more predictor variables and a .

If we only have one predictor variable and one response variable, we can use , which uses the following formula to estimate the relationship between the variables:

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

where:

y : The estimated response value.

β_0 : The average value of y when x is zero.

β_1 : The average change in y associated with a one unit increase in x .

x : The value of the predictor variable.

Simple linear regression uses the following null and alternative hypotheses:

$$H_0: \beta_1 = 0$$

$$H_A: \beta_1 \neq 0$$

The null hypothesis states that the coefficient β_1 is equal to zero. In other words, there is no statistically significant relationship between the predictor variable, x , and the response variable, y .

The alternative hypothesis states that β_1 is *not* equal to zero. In other words, there *is* a statistically significant relationship between x and y .

If we have multiple predictor variables and one response variable, we can use , which uses the following formula to estimate the relationship between the variables:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

where:

y : The estimated response value.

β_0 : The average value of y when all predictor variables are equal to zero.

β_i : The average change in y associated with a one unit increase in x_i .

x_i : The value of the predictor variable x_i .

Multiple linear regression uses the following null and alternative hypotheses:

H₀: $\beta_1 = \beta_2 = \dots = \beta_k = 0$

H_A: $\beta_1 = \beta_2 = \dots = \beta_k \neq 0$

The null hypothesis states that all coefficients in the model are equal to zero. In other words, none of the predictor variables have a statistically significant relationship with the response variable, y .

The alternative hypothesis states that not every coefficient is simultaneously equal to zero.

Example 1: Simple Linear Regression

Suppose a professor would like to use the number of hours studied to predict the exam score that students will receive in his class. He collects data for 20 students and fits a simple linear regression model.

The following screenshot shows the output of the regression model:

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

The fitted simple linear regression model is:

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

To determine if there is a statistically significant relationship between hours studied and exam score, we need to analyze the of the model and the corresponding p-value:

Overall F-Value: **47.9952**

P-value: **0.000**

Since this p-value is less than .05, we can reject the null hypothesis. In other words, there is a statistically significant relationship between hours studied and exam score received.

Example 2: Multiple Linear Regression

Suppose a professor would like to use the number of hours studied and the number of prep exams taken to predict the exam score that students will receive in his class. He collects data for 20 students and fits a multiple linear regression model.

The following screenshot shows the output of the regression model:

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 fitted multiple linear regression model is:

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

To determine if there is a jointly statistically significant relationship between the two predictor variables and the response variable, we need to analyze the overall F value of the model and the corresponding p-value:

Overall F-Value: **23.46**

P-value: **0.00**

Since this p-value is less than .05, we can reject the null hypothesis. In other words, hours studied and prep exams taken have a jointly statistically significant relationship with exam score.

Note: Although the p-value for prep exams taken ($p = 0.52$) is not significant, prep exams *combined* with hours studied has a significant relationship with exam score.

ARABPSYCHOLOGY.COM