

How can I perform multiple linear regression by hand using a step-by-step approach?

Authored by
stats writer

April 22, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I perform multiple linear regression by hand using a step-by-step approach?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=138058>

Multiple linear regression is a statistical technique used to analyze the relationship between multiple independent variables and a single dependent variable. It is commonly used in various fields, such as economics, psychology, and social sciences, to predict and explain the behavior of a dependent variable based on the values of the independent variables.

Performing multiple linear regression by hand using a step-by-step approach involves several key steps. First, the data must be collected and organized into a matrix format. Then, the regression coefficients can be calculated using the least squares method, which involves minimizing the sum of the squared differences between the actual and predicted values of the dependent variable. This process can be done manually using mathematical equations and calculations.

Next, the significance of the regression coefficients can be determined by calculating the t-statistic and comparing it to a critical value. This helps to identify which independent variables are statistically significant in predicting the dependent variable.

Furthermore, the overall fit of the regression model can be assessed by calculating the coefficient of determination (R-squared) and the adjusted R-squared. These measures indicate the proportion of variation in the dependent variable that can be explained by the independent variables.

Finally, the results of the analysis can be interpreted and used to make predictions and draw conclusions about the relationship between the variables.

In summary, performing multiple linear regression by hand using a step-by-step approach requires careful data collection and organization, as well as the use of mathematical calculations and statistical tests to determine the significance and fit of the regression model. This process can provide valuable insights and predictions in various research and practical applications.

Multiple Linear Regression by Hand (Step-by-Step)

Multiple linear regression is a method we can use to quantify the relationship between two or more predictor variables and a response variable.

This tutorial explains how to perform multiple linear regression by hand.

Example: Multiple Linear Regression by Hand

Suppose we have the following dataset with one response variable y and two predictor variables X_1 and X_2 :

y	X_1	X_2
140	60	22
155	62	25
159	67	24
179	70	20
192	71	15
200	72	14
212	75	14
215	78	11

Use the following steps to fit a multiple linear regression model to this dataset.

Step 1: Calculate X_1^2 , X_2^2 , X_1y , X_2y and X_1X_2 .

	y	X_1	X_2
	140	60	22
	155	62	25
	159	67	24
	179	70	20
	192	71	15
	200	72	14
	212	75	14
	215	78	11
Mean	181.5	69.375	18.125
Sum	1452	555	145

	X_1^2	X_2^2	X_1y	X_2y	X_1X_2
	3600	484	8400	3080	1320
	3844	625	9610	3875	1550
	4489	576	10653	3816	1608
	4900	400	12530	3580	1400
	5041	225	13632	2880	1065
	5184	196	14400	2800	1008
	5625	196	15900	2968	1050
	6084	121	16770	2365	858
Sum	38767	2823	101895	25364	9859

Step 2: Calculate Regression Sums.

Next, make the following regression sum calculations:

$$\begin{aligned}\Sigma x_1^2 &= \Sigma X_1^2 - (\Sigma X_1)^2 / n = 38,767 - (555)^2 / 8 = 263.875 \\ \Sigma x_2^2 &= \Sigma X_2^2 - (\Sigma X_2)^2 / n = 2,823 - (145)^2 / 8 = 194.875 \\ \Sigma x_1y &= \Sigma X_1y - (\Sigma X_1 \Sigma y) / n = 101,895 - (555 * 1,452) / 8 = 1,162.5 \\ \Sigma x_2y &= \Sigma X_2y - (\Sigma X_2 \Sigma y) / n = 25,364 - (145 * 1,452) / 8 = -953.5 \\ \Sigma x_1x_2 &= \Sigma X_1X_2 - (\Sigma X_1 \Sigma X_2) / n = 9,859 - (555 * 145) / 8 = -200.375\end{aligned}$$

	y	X ₁	X ₂		X ₁ ²	X ₂ ²	X ₁ y	X ₂ y	X ₁ X ₂
	140	60	22		3600	484	8400	3080	1320
	155	62	25		3844	625	9610	3875	1550
	159	67	24		4489	576	10653	3816	1608
	179	70	20		4900	400	12530	3580	1400
	192	71	15		5041	225	13632	2880	1065
	200	72	14		5184	196	14400	2800	1008
	212	75	14		5625	196	15900	2968	1050
	215	78	11		6084	121	16770	2365	858
Mean	181.5	69.375	18.125	Sum	38767	2823	101895	25364	9859
Sum	1452	555	145						

Reg Sums	263.875	194.875	1162.5	-953.5	-200.375

Step 3: Calculate b0, b1, and b2.

The formula to calculate b1 is: /

Thus, b1 = / = 3.148

The formula to calculate b_2 is: /

Thus, $b_2 = / = -1.656$

Thus, $b_0 = 181.5 - 3.148(69.375) - (-1.656)(18.125) = -6.867$

Step 5: Place b_0 , b_1 , and b_2 in the estimated linear regression equation.

The estimated linear regression equation is: $\hat{y} = b_0 + b_1x_1 + b_2x_2$

In our example, it is $\hat{y} = -6.867 + 3.148x_1 - 1.656x_2$

How to Interpret a Multiple Linear Regression Equation

Here is how to interpret this estimated linear regression equation: $\hat{y} = -6.867 + 3.148x_1 - 1.656x_2$

$b_0 = -6.867$. When both predictor variables are equal to zero, the mean value for y is -6.867 .

$b_1 = 3.148$. A one unit increase in x_1 is associated with a 3.148 unit increase in y , on average, assuming x_2 is held constant.

$b_2 = -1.656$. A one unit increase in x_2 is associated with a 1.656 unit decrease in y , on average, assuming x_1 is held constant.

An Introduction to Multiple Linear Regression

How to Perform Simple Linear Regression by Hand

ARABPSYCHOLOGY.COM