Can you explain how to interpret P-values in linear regression, including an example?

Authored by

stats writer

June 28, 2024

P-values in linear regression are a statistical measure used to determine the significance of the relationship between the independent and dependent variables in a regression model. It is a measure of the probability that the observed results are due to chance, rather than a true relationship between the variables. A low p-value (typically less than 0.05) indicates that there is a significant relationship between the variables, while a high p-value suggests that the relationship is not significant.

For example, if we are trying to determine the effect of education level on income, we would use a linear regression model. The P-value would tell us the likelihood of the observed relationship between education level and income being due to chance. If the p-value is low, it means that there is a significant relationship between education level and income, and we can conclude that education level does have an impact on income. On the other hand, if the p-value is high, it indicates that the relationship between education level and income is not significant, and we cannot conclude that education level has a direct effect on income.

Interpret P-Values in Linear Regression (With Example)

In statistics, linear regression models are used to quantify the relationship between one or more predictor variables and a .

Whenever you perform regression analysis using some statistical software, you will receive a regression table that summarizes the results of the model.

Two of the most important values in a regression table are the regression coefficients and their corresponding p-values.

The p-values tell you whether or not there is a statistically significant relationship between each

predictor variable and the response variable.

The following example shows how to interpret the p-values of a model in practice.

Example: Interpreting P-Values in Regression Model

Suppose we want to fit a regression model using the following variables:

Predictor Variables

Total number of hours studied (between 0 and 20)Whether or not a student used a tutor (yes or no)

Response Variable

Exam score (between 0 and 100)

We want to examine the relationship between the predictor variables and the response variable to find out if hours studied and tutoring actually have a meaningful impact on exam score.

Suppose we run a regression analysis and get the following output:

Term	Coefficient	Standard Error	t Stat	P-value

Intercept	48.56	14.32	3.39	0.002
Hours studied	2.03	0.67	3.03	0.009
Tutor	8.34	5.68	1.47	0.138

Here's how to interpret the output for each term in the model:

Interpreting the P-value for Intercept

The intercept term in a regression table tells us the average expected value for the response variable when all of the predictor variables are equal to zero.

The p-value is 0.002, which tells us that the intercept term is statistically different than zero.

In practice, we don't usually care about the p-value for the intercept term. Even if the p-value isn't less than some significance level (e.g. 0.05), we would still keep the intercept term in the model.

Interpreting the P-value for a Continuous Predictor Variable

In this example, Hours studied is a continuous predictor variable that ranges from 0 to 20 hours.

From the regression output, we can see that the regression coefficient for Hours studied is 2.03. This

means that, on average, each additional hour studied is associated with an increase of 2.03 points on the final exam, assuming the predictor variable Tutor is held constant.

For example, consider student A who studies for 10 hours and uses a tutor. Also consider student B who studies for 11 hours and also uses a tutor. According to our regression output, student B is expected to receive an exam score that is 2.03 points higher than student A.

The corresponding p-value is 0.009, which is statistically significant at an alpha level of 0.05.

This tells us that that the average change in exam score for each additional hour studied is statistically significantly different than zero.

Another way to put this: Hours studied has a statistically significant relationship with the response variable exam score.

Interpreting the P-value for a Categorical Predictor Variable

In this example, Tutor is a categorical predictor variable that can take on two different values:

1 = the student used a tutor to prepare for the exam0 = the student did not used a tutor to prepare for the exam

From the regression output, we can see that the regression coefficient for Tutor is 8.34. This means that, on average, a student who used a tutor scored 8.34 points higher on the exam compared to a student who did not used a tutor, assuming the predictor variable Hours studied is held constant.

For example, consider student A who studies for 10 hours and uses a tutor. Also consider student B who studies for 10 hours and does not use a tutor. According to our regression output, student A is expected to receive an exam score that is 8.34 points higher than student B.

The corresponding p-value is 0.138, which is not statistically significant at an alpha level of 0.05.

This tells us that that the average change in exam score for each additional hour studied is not statistically significantly different than zero.

Another way to put this: The predictor variable Tutor

does not have a statistically significant relationship with the response variable exam score.

This indicates that although students who used a tutor scored higher on the exam, this difference could have been due to random chance.

Additional Resources

The following tutorials provide additional information about linear regression: