

What is the interpretation of adjusted R-squared and can you provide examples?

Authored by
stats writer

June 29, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the interpretation of adjusted R-squared and can you provide examples?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=158443>

Adjusted R-squared is a statistical measure that assesses the goodness of fit of a regression model. It takes into account the number of independent variables in the model and adjusts the R-squared value accordingly. This helps to avoid overestimation of the model's predictive power by penalizing the addition of unnecessary variables.

For example, let's say a researcher is trying to predict a person's height using their age and weight as independent variables. The R-squared value for this model may be 0.85, indicating a good fit. However, when the researcher adds a third independent variable, such as shoe size, the R-squared value may increase to 0.90. This increase may be misleading as shoe size is not a relevant predictor of height. Adjusted R-squared takes into account the addition of this unnecessary variable and adjusts the value to a more accurate representation of the model's predictive power.

In summary, adjusted R-squared provides a more reliable interpretation of a regression model's goodness of fit by penalizing the inclusion of irrelevant variables. It is a useful tool for evaluating the effectiveness of a model in predicting outcomes.

Interpret Adjusted R-Squared (With Examples)

When we fit we often calculate the R-squared value of the model.

The R-squared value is the proportion of the variance in the response variable that can be explained by the predictor variables in the model.

The value for R-squared can range from 0 to 1 where:

A value of 0 indicates that the response variable cannot be explained by the predictor variables at all. A value of 1 indicates that the response variable can be perfectly explained by the predictor variables.

Although this metric is commonly used to assess how well a regression model fits a dataset, it has one serious drawback:

The drawback of R-squared:

R-squared will always increase when a new predictor variable is added to the regression model.

Even if a new predictor variable is almost completely unrelated to the response variable, the R-squared value of the model will increase, if only by a small amount.

For this reason, it's possible that a regression model with a large number of predictor variables has a high R-squared value, even if the model doesn't fit the data well.

Fortunately there is an alternative to R-squared known as adjusted R-squared.

The adjusted R-squared is a modified version of R-squared that adjusts for the number of predictors in a regression model.

It is calculated as:

Adjusted R² = 1 -

where:

R²: The R² of the model
n: The number of observations
k: The number of predictor variables

Because R-squared always increases as you add more predictors to a model, the adjusted R-squared can tell you how useful a model is, *adjusted for the number of predictors in a model.*

The advantage of Adjusted R-squared:

Adjusted R-squared tells us how well a set of predictor variables is able to explain the variation in the response variable, *adjusted for the number of predictors in a model.*

Because of the way it's calculated, adjusted R-squared can be used to compare the fit of regression models with different numbers of predictor variables.

To gain a better understanding of adjusted R-squared, check out the following example.

Example: Understanding Adjusted R-Squared in Regression Models

Suppose a professor collects data on students in his class and fits the following regression model to understand how hours spent studying and current grade in the class affect the score a student receives on the final exam.

$$\text{Exam Score} = \beta_0 + \beta_1(\text{hours spent studying}) + \beta_2(\text{current grade})$$

Suppose this regression model has the following metrics:

R-squared: 0.955 Adjusted R-squared: 0.946

Now suppose the professor decides to collect data on another variable for each student: shoe size.

Although this variable should be completely unrelated to the final exam score, he decides to fit the following regression model:

$$\text{Exam Score} = \beta_0 + \beta_1(\text{hours spent studying}) + \beta_2(\text{current grade}) + \beta_3(\text{shoe size})$$

Suppose this regression model has the following metrics:

R-squared: 0.965 Adjusted R-squared: 0.902

If we only looked at the R-squared values for each of these two regression models, we would conclude that the second model is better to use because it has a higher R-squared value!

However, if we look at the adjusted R-squared values then we come to a different conclusion: The first model is better to use because it has a higher adjusted R-squared value.

The second model only has a higher R-squared value because it has more predictor variables than the first model.

However, the predictor variable that we added (shoe size) was a poor predictor of final exam score, so the adjusted R-squared value penalized the model for adding this predictor variable.

This example illustrates why adjusted R-squared is a better metric to use when comparing the fit of

regression models with different numbers of predictor variables.

Additional Resources

The following tutorials explain how to calculate adjusted R-squared values using different statistical software:

ARABPSYCHOLOGY.COM