How to Easily Understand the Difference Between R and R-Squared

Authored by
stats writer

December 5, 2025

In the world of statistics and data analysis, few concepts are as fundamental yet frequently confused as **R** and **R-squared**, often written as R2. While mathematically related, these two metrics serve distinct purposes when evaluating the strength and utility of a statistical model, particularly in regression analysis. Understanding the precise definition and interpretation of each is critical for drawing valid conclusions from your data.

Distinguishing R and R-Squared: The Core Concepts

The confusion often stems from the fact that both R and R-squared quantify relationship strength, but they measure fundamentally different aspects of the relationship. R, or the correlation coefficient, describes the direction and linearity of the relationship between variables. In contrast, R-squared, or the Coefficient of Determination, measures the predictive power of the model itself by quantifying how much variation in the dependent variable is accounted for by the independent variable(s).

To establish a clear framework, let us define these terms based on their primary applications:

R: The Correlation Coefficient

The term **R** is the Pearson correlation coefficient, which measures the linear association between two variables. Its value always falls between -1 and 1. A value of 1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, and 0 indicates no linear correlation.

R in Bivariate Analysis: Represents the correlation between the predictor variable (x) and the response variable (y).
R in Regression Context: Represents the correlation between the observed values of the response variable and the specific values of the response variable predicted by the model.

In essence, R provides immediate insight into the strength and directionality of the linear trend observed in the data.

R-Squared: Quantifying Explained Variance

The term **R2**, or the Coefficient of Determination, measures the proportion of the total variability in the response variable that is explained by the regression model. It is a powerful metric for assessing the model's goodness-of-fit.

R2 Interpretation: This metric represents the proportion of the total variance in the response variable (y) that can be explained by the predictor variables included in the model.

The value for R2 ranges strictly between 0 and 1. The closer the value is to 1 (or 100%), the stronger the collective relationship between the predictor variable(s) and the response variable, indicating that the model accounts for most of the observed fluctuation in the outcome.

The Relationship Between R and R-Squared

The clear mathematical relationship between R and R2 is most apparent when dealing with Simple Linear Regression (SLR). In SLR, since there is only one predictor, R2 is simply the square of R. This squaring operation automatically ensures R2 is non-negative and removes the directional information provided by the sign of R, focusing purely on explanatory magnitude.

This simple squaring rule, however, does not directly apply to individual predictor correlations in Multiple Linear Regression (MLR). In MLR, R2 is calculated based on the combined explanatory power of all predictors, while R is calculated based on the correlation between the observed outcome and the outcome predicted by the full model.

Example 1: Applying Simple Linear Regression

Let us examine how these metrics are interpreted using a simple case study involving a single predictor. Suppose we analyze a small dataset detailing the hours studied and the subsequent exam score received by 12 students in a mathematics course:

Study Hours	Exam Score
1	58
1	61
2	62
2	65
1	65
2	68
2	72
3	74
3	78
4	85
4	90
5	95

Using statistical software (such as R, Python, or Excel), we fit a simple linear regression model where "study hours" serves as the single predictor variable and "exam score" is the response variable. The resulting statistical output provides the necessary metrics:

Regression Statistics	
R	0.959
R Square	0.920
Adjusted R Square	0.912
Standard Error	3.566
Observations	12

Interpretation of Simple Linear Regression Results

The output provides clear values for both R and R2:

R (Correlation): The correlation coefficient between hours studied and exam score is determined to be **0.959**. This signifies a very strong, positive linear relationship, confirming that greater study time is closely associated with higher scores.
R2 (Explained Variance): The R-squared value for this regression model is **0.920**. This tells us that 92.0% of the total variation in the exam scores can be effectively explained by the variation in the number of hours studied.

As expected in simple linear regression, the R2 value is precisely the square of the R value:

R2 = R * R = 0.959 * 0.959 = **0.920**

Example 2: Exploring Multiple Linear Regression

Next, we introduce a second predictor variable to see how the model metrics evolve. We now include the student's "Current Grade" alongside "Hours Studied" to predict the "Exam Score." This scenario requires multiple linear regression:

Study Hours	Current Grade	Exam Score
1	65	58
1	78	61
2	76	62
2	76	65
1	79	65
2	80	68
2	81	72
3	84	74
3	88	78
4	85	85
4	96	90
5	90	95

We fit the multiple linear regression model using both "study hours" and "current grade" as the collective predictor set. The goal is to see if adding prior academic performance improves the explanation of the final exam score variability. The resulting statistical output is as follows:

Regression Statistics	
Multiple R	0.978
R Square	0.956
Adjusted R Square	0.946
Standard Error	2.790
Observations	12

Interpretation of Multiple Linear Regression Results

The interpretation of R and R2 provides insight into the model's performance when utilizing multiple variables:

R (Multiple Correlation Coefficient): In this MLR context, R represents the correlation between the actual exam scores and the set of scores predicted by the complex model incorporating both predictors. The value is **0.978**, indicating that the predicted scores are extremely closely aligned with the actual outcomes.

R2 (Explained Variance): The R-squared for this expanded regression model is **0.956**. This demonstrates that 95.6% of the variation in the exam scores can now be explained jointly by the number of hours studied and the student's current grade in the class. This higher R2 confirms that the second predictor added significant explanatory power.

In line with the fundamental definition, the R2 value is equal to the square of the Multiple Correlation Coefficient (R):

R2 = R * R = 0.978 * 0.978 = **0.956**

Summary of Key Differences

While frequently conflated, R and R-squared provide complementary but distinct information regarding the suitability and fit of a regression model. Choosing the appropriate metric for reporting depends on whether you seek to understand the simple linear strength (R) or the overall explanatory power (R2).

R (Correlation Coefficient): Measures the strength and direction of linear association, ranging from -1 to 1. In regression, it is used to assess how well the model predictions correlate with the actual data.
R2 (Coefficient of Determination): Measures the proportion of the total variance in the dependent variable explained by the entire set of independent variables, ranging from 0 to 1. It is the primary metric for assessing overall model fit.

Ultimately, R2 is generally a more informative measure for evaluating the predictive utility of a regression model, as it directly quantifies the percentage of the outcome variability accounted for by the inputs.