

What's the Difference Between Multiple R and R-Squared?

Authored by
stats writer

December 9, 2025

RECOMMENDED CITATION

stats writer (2025). *What's the Difference Between Multiple R and R-Squared?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=106885>

In the realm of statistics and linear regression modeling, researchers frequently encounter metrics designed to evaluate the strength and usefulness of their predictive models. Among the most common are Multiple R and R-Squared. While these two statistics are fundamentally related—one being derived mathematically from the other—they provide distinctly different insights into model performance. Understanding this difference is crucial for accurately interpreting the output of any regression analysis.

Multiple R quantifies the strength of the linear relationship between the observed values and the values predicted by the model. It is, essentially, the simple correlation coefficient between these two sets of data points. Conversely, R-squared, often called the Coefficient of Determination, measures the proportion of the variance in the dependent variable that can be statistically accounted for by the independent variables included in the model. This detailed exploration will clarify the interpretation, calculation, and practical application of both metrics, along with the necessary refinement provided by Adjusted R-Squared.

Understanding the Multiple Correlation Coefficient (Multiple R)

When running a multiple linear regression analysis using standard statistical software packages, one of the primary outputs you will observe is the value labeled **Multiple R**. This metric serves as a robust indicator of the quality of fit for the regression line relative to the actual data points. More precisely, **Multiple R** measures the correlation between the outcome variable (Y) and the set of predictor variables (X) combined, as represented by the predicted outcome (\hat{Y}).

In simple terms, **Multiple R** is calculated as the Pearson correlation coefficient between the observed values of the dependent variable and the values predicted by your constructed regression equation. This value is always non-negative, ranging from 0 to 1. A value approaching 1 signifies a very strong linear relationship, indicating that the model's predictions closely align with the actual observed data. Conversely, a value near 0 suggests a weak or non-existent linear relationship, meaning the model is poor at predicting the outcome.

It is important to recognize that **Multiple R** is primarily a measure of relationship strength, conceptually similar to the simple correlation coefficient (r) found in bivariate analysis, but extended to account for multiple predictors simultaneously. While it confirms the overall fitness of the model predictions, it does not directly tell us the proportion of the dependent variable's variability that is explained, which is the specialized role reserved for R-Squared.

Interpreting the Coefficient of Determination (R-Squared)

The metric known as **R-Squared** (or R^2) is arguably the most frequently cited measure of model performance in regression analysis. Unlike Multiple R, which focuses on correlation strength, R-

squared directly addresses explanatory power. It quantifies the proportion of the total variation in the response (dependent) variable that is successfully explained by the independent (predictor) variables included in the linear regression model.

Mathematically, **R-Squared** is calculated by simply squaring the **Multiple R** value. This relationship means that if Multiple R is 0.9, R-Squared is 0.81 (or 81%). This interpretation is highly intuitive: an R-Squared of 0.81 suggests that 81% of the total variance observed in the dependent variable can be attributed to the effects captured by the predictor variables in the model. The remaining 19% of the variance is unexplained, usually attributed to random error or factors not included in the analysis.

Because **R-Squared** is inherently tied to the total sum of squares and the residual sum of squares, its value must always fall within the interval of 0 and 1 (or 0% and 100%). A value of 0 indicates that the model explains none of the variability of the response data around its mean, essentially meaning the predictors are useless. A value of 1 implies that the model perfectly predicts the response values, a scenario rarely achieved with real-world data. Researchers prioritize R-Squared because it provides a direct, interpretable measure of the model's overall predictive utility.

The Flaw: Why Standard R-Squared Can Be Misleading

Despite its usefulness, standard **R-Squared** suffers from a significant systematic bias when applied to models with numerous predictors. A critical limitation of this statistic is that it is guaranteed to increase—or at minimum, remain constant—every single time an additional independent variable is added to the regression model, regardless of whether that new variable is statistically significant or contributes meaningfully to the explanatory power.

This inflation occurs because adding any new variable, even pure noise, allows the model to capture a minute portion of the remaining residual variance, marginally reducing the error sum of squares. Consequently, if a researcher continuously adds irrelevant predictors, the R-Squared value will climb higher, potentially leading to the false conclusion that a complex, overfit model is superior to a simpler, more parsimonious one. This risk is particularly high when dealing with small sample sizes or large numbers of predictors.

To mitigate this inherent bias and penalize model complexity that does not yield genuine statistical improvement, statisticians rely on a refined metric: the **Adjusted R-Squared**. This adaptation provides a more honest assessment of model fitness, crucial for selecting the best set of predictors.

The Refinement: How Adjusted R-Squared Corrects for Model Complexity

The **adjusted R-squared** is a modified and substantially more reliable version of standard R-

Squared. Its primary purpose is to account for the number of predictor variables included in the linear regression model, thereby penalizing the addition of irrelevant variables. This penalty mechanism ensures that the value only increases if the new predictor improves the model fit more than would be expected by chance.

Unlike R-Squared, the value of **Adjusted R-Squared** can actually decrease when a predictor variable is added that does not significantly contribute to explaining the dependent variable's variance. This crucial feature makes it the preferred metric when comparing models that utilize different numbers of independent variables, especially during the crucial process of variable selection or model refinement.

The calculation of **Adjusted R-Squared** incorporates degrees of freedom into the equation, providing the necessary adjustment for model size. It is calculated using the following formula:

Adjusted R² = 1 -

In this formula, the terms represent essential components of the model structure:

R²: The standard R-Squared of the model, representing explained proportion of variance.

n: The total number of observations (data points) used in the analysis.

k: The number of predictor variables (independent variables) included in the equation.

Because it integrates this penalty for complexity, **Adjusted R-Squared** is superior for model comparison. It helps analysts identify the most parsimonious and effective model, ensuring that explanatory power is achieved without overfitting the data.

Comparing Multiple R, R-Squared, and Adjusted R-Squared in Practice

To summarize the utility of these three statistical measures, it is helpful to delineate their primary roles in the context of statistical modeling. While all three are related by calculation, their interpretations serve distinct analytical purposes.

Multiple R tells us about the direction and magnitude of the linear correlation coefficient between the actual outcomes and the predicted outcomes. It is a metric of how tightly the predicted values cluster around the observed values, providing a measure of the linear strength. However, this metric is less frequently cited in final reports because it does not offer the direct explanatory percentage that R-Squared provides.

R-Squared offers the most intuitive interpretation: the percentage of total variability explained by the model. It is a powerful tool for initial assessment of explanatory success. Analysts often use this value initially to determine if the predictors collectively hold value in explaining the dependent variable. Nevertheless, due to its propensity for inflation when complexity increases, it is unreliable

for model selection.

The **Adjusted R-Squared** is the gold standard for model comparison. Whenever a researcher needs to choose between Model A (with three predictors) and Model B (with five predictors), relying solely on the standard R-Squared will favor the more complex model. By using the adjusted statistic, the researcher can accurately determine which model achieves the highest explanatory power relative to its degrees of freedom, thus favoring the simplest robust model.

To gain a concrete, practical understanding of how these metrics interact and differ, we will now examine a detailed example derived from a student performance dataset.

Example: Calculating and Interpreting Multiple R, R-Squared, & Adjusted R-Squared

Suppose we are interested in predicting student performance based on preparatory effort and prior academic achievement. We have collected a dataset containing the following three variables for 12 different students ($n=12$): Study Hours, Current Grade (as a proxy for prior achievement), and the final Exam Score.

The structure of our raw data is shown below:

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 decide to fit a multiple linear regression model using *Study Hours* and *Current Grade* as the two predictor variables ($k=2$) and *Exam Score* as the response variable. After running the analysis in our statistical software, we obtain the following summary output, which provides the critical fitness statistics:

SUMMARY OUTPUT

<i>Regression Statistics</i>					
Multiple R		0.978			
R Square		0.956			
Adjusted R Square		0.946			
Standard Error		2.790			
Observations		12			

<i>ANOVA</i>					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>p-value</i>
Regression	2	1516.192	758.096	97.388	0.000
Residual	9	70.058	7.784		
Total	11	1586.250			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	17.175	12.556	1.368	0.205
Study Hours	6.384	1.087	5.874	0.000
Current Grade	0.486	0.179	2.709	0.024

Based on this output, we can systematically analyze the values obtained for the three key model metrics: Multiple R, R-Squared, and Adjusted R-Squared.

Analyzing the Model Output Metrics

Multiple R: 0.978. This value represents the correlation coefficient between the observed exam scores and the exam scores predicted by the regression equation incorporating both study hours and current grade. A value very close to 1 indicates an extremely strong linear relationship between the combined predictors and the outcome variable, suggesting high predictive accuracy by the model.

R Square (R²): 0.956. This calculation confirms the mathematical relationship: $(\text{Multiple R})^2 = (0.978)^2 \approx 0.956$. This means that 95.6% of the total variation in the Exam Scores across the 12 students can be successfully explained by the combined influence of the number of hours spent studying and the student's current grade. This leaves only 4.4% of the variability unexplained, reinforcing that this model is an excellent fit for the data.

Adjusted R-Square: 0.946. This value incorporates the sample size ($n=12$) and the number of predictors ($k=2$) to adjust the R-Squared value for complexity. The calculation follows the defined formula precisely:

Adjusted R² = 1 -

$$= 1 - 1 - \approx 0.946.$$

The difference between the raw R-Squared (0.956) and the **Adjusted R-Squared** (0.946) is relatively small in this example, which is typical for models with few predictors and a decent sample size. However, this adjusted metric provides a critical benchmark for comparison against alternative models.

Using Adjusted R-Squared for Robust Model Selection

The true utility of the **Adjusted R-Squared** becomes apparent when we engage in model comparison. Imagine we were to fit a second regression model (Model C) that attempts to predict Exam Score using ten different variables ($k=10$), perhaps including height, favorite color, and shoe size, alongside Study Hours and Current Grade. Due to the inherent inflation mechanism, Model C's raw R-Squared might increase slightly, perhaps reaching 0.965.

If we were to calculate the Adjusted R-Squared for this unnecessarily complex Model C, we might find that the value drops significantly, perhaps to **0.88**. This substantial decrease, relative to our initial two-predictor model's 0.946, decisively indicates that the additional eight predictors are adding little explanatory power while introducing unwanted complexity and reducing degrees of freedom.

Therefore, the adjusted metric allows us to conclude that the regression model with just two robust predictors (Study Hours and Current Grade) is demonstrably better than the over-complicated model with ten predictors because it maintains a higher explanatory proportion when adjusted for complexity. In essence, **Adjusted R-Squared** drives the analyst toward the principles of parsimony—finding the simplest model that explains the maximum amount of variation.