

R Squared and Adjusted R Squared Explained

R-squared is one of the most widely reported and most widely misread regression statistics. This page explains what it measures, what adjusted R-squared corrects for, and what neither one guarantees.



R Squared and Adjusted R Squared Explained

Exported from statreason.com. Deterministic, browser-local content; not professional statistical or research advice.

What this answers

This page answers "what does an R-squared of 0.65 actually mean, and is a higher number always better?" R-squared describes the proportion of variance in the outcome that the fitted model accounts for within your observed sample. It is a fit statistic, not a measure of prediction accuracy on new data or a signal that the model's predictors cause the outcome.

What R-squared measures

R-squared is the proportion of total variation in the outcome variable that the model's predictors explain, ranging from 0, meaning the model explains none of the variation, to 1, meaning the model fits the observed data perfectly. An R-squared of 0.65 means 65 percent of the variation in the outcome, within this specific sample, is associated with the predictors in the model; the remaining 35 percent is unexplained by this model.

Why adjusted R-squared exists

Plain R-squared can only increase, or stay the same, every time you add another predictor to a model, even a predictor with no real relationship to the outcome, because it can only ever capture more in-sample variation by chance. Adjusted R-squared applies a penalty based on the number of predictors relative to the sample size, so it can decrease when an added predictor does not meaningfully improve the fit, which makes it more useful for comparing models with different numbers of predictors.

What neither one proves

A high R-squared does not mean the model's predictors cause the outcome, does not guarantee the model will predict new, unseen data well, and does not mean every individual prediction will be close to accurate. A model with a modest R-squared can still be useful for planning purposes, and a model with a very high R-squared can still perform poorly on new data if it was overfit to the specific sample it was built on.

Worked example

A model predicting exam score from hours studied has an R-squared of 0.40. Adding a predictor for the student's favorite color, which has no real relationship to exam performance, still raises plain R-squared slightly, purely by chance fit to this specific sample. Adjusted R-squared for that same addition is likely to fall, correctly signaling that the new predictor did not earn its place in the model.

Assumption audit

Calculated from your data: both R-squared and adjusted R-squared, automatically, whenever you fit a model with the linked regression calculators.

Evidence to review: a residual plot alongside R-squared, since a high R-squared with a clearly patterned residual plot still signals a misspecified model.

You must verify: what R-squared value is actually meaningful in your field, since acceptable values vary enormously between, for example, controlled laboratory measurements and noisy human-behavior data.

Source

This explanation follows the standard treatment of R-squared and adjusted R-squared in the NIST/SEMATECH e-Handbook of Statistical Methods and the shared statistical reasoning contract every StatReason engine is built against.

Limitations

This page covers R-squared for linear regression models; a pseudo-R-squared reported for logistic regression is calculated differently and does not have the same variance-explained

interpretation.

Next action: compare model fit using the [Adjusted R-Squared Calculator](#), or fit a new model with the [Multiple Linear Regression Calculator](#).