WIKIPEDIA

Coefficient of determination

In statistics, the **coefficient of determination**, denoted R^2 or r^2 and pronounced "R squared", is the proportion of the variation in the dependent variable that is predictable from the independent variable(s).
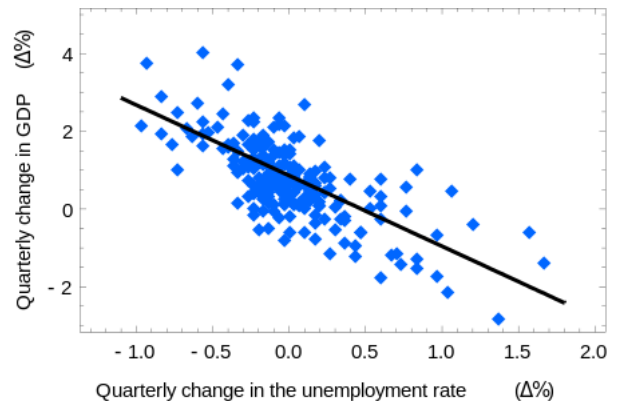
It is a statistic used in the context of statistical models whose main purpose is either the prediction of future outcomes or the testing of hypotheses, on the basis of other related information. It provides a measure of how well observed outcomes are replicated by the model, based on the proportion of total variation of outcomes explained by the model.[1][2][3]

There are several definitions of R^2 that are only sometimes equivalent. One class of such cases includes that of simple linear regression where r^2 is used instead of R^2. When only an intercept is included, then r^2 is simply the square of the sample correlation coefficient (i.e., r) between the observed outcomes and the observed predictor values.[4] If additional regressors are included, R^2 is the square of the coefficient of multiple correlation. In both such cases, the coefficient of determination normally ranges from 0 to 1.

There are cases where the computational definition of R^2 can yield negative values, depending on the definition used. This can arise when the predictions that are being compared to the corresponding outcomes have not been derived from a model-fitting procedure using those data. Even if a model-fitting procedure has been used, R^2 may still be negative, for example when linear regression is conducted without including an intercept,[5] or when a non-linear function is used to fit the data.[6] In cases where negative values arise, the mean of the data provides a better fit to the outcomes than do the fitted function values, according to this particular criterion.

Ordinary least squares regression of Okun's law. Since the regression line does not miss any of the points by very much, the R^2 of the regression is relatively high.
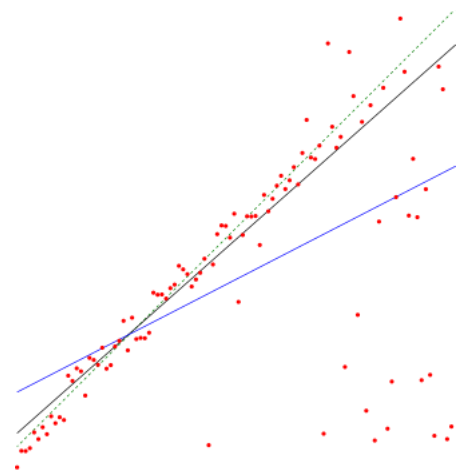
Comparison of the Theil–Sen estimator (black) and simple linear regression (blue) for a set of points with outliers. Because of the many outliers, neither of the regression lines fits the data well, as measured by the fact that neither gives a very high R^2.

An article published in the PeerJ Computer Science journal in 2021 states that the coefficient of determination can be more truthful than SMAPE, MAE, MAPE, MSE, and RMSE in regression analysis evaluation.[7]

When evaluating the goodness-of-fit of simulated (Y_{pred}) vs. measured (Y_{obs}) values, it is not appropriate to base this on the R^2 of the linear regression (i.e., $Y_{\text{obs}} = m \cdot Y_{\text{pred}} + b$). The R^2 quantifies the degree of any linear correlation between Y_{obs} and Y_{pred}, while for the goodness-of-fit evaluation only one specific linear correlation should be taken into consideration: $Y_{\text{obs}} = 1 \cdot Y_{\text{pred}} + 0$ (i.e., the 1:1 line).[8][9]

Contents

Definitions

A [data set](#) has n values marked $y_1,...,y_n$ (collectively known as y_i or as a vector $\boldsymbol{y} = [y_1,...,y_n]^T$), each associated with a fitted (or modeled, or predicted) value $f_1,...,f_n$ (known as f_i, or sometimes $\hat{y}_i$, as a vector $\boldsymbol{f}$).

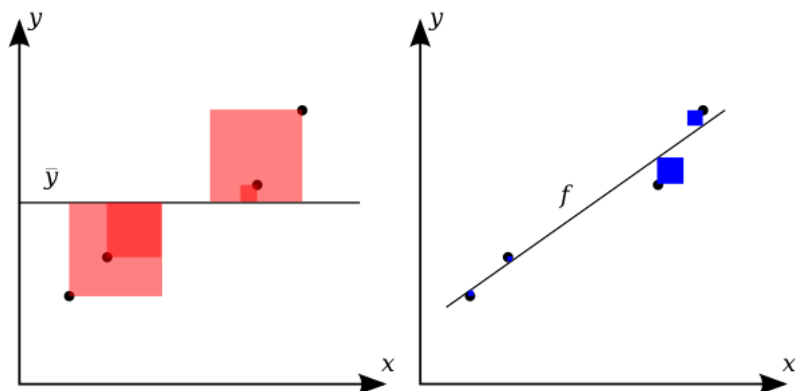Define the [residuals](#) as $e_i = y_i - f_i$ (forming a vector $\boldsymbol{e}$).

If $\bar{y}$ is the mean of the observed data:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

then the variability of the data set can be measured with two [sums of squares](#) formulas:

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

The better the linear regression (on the right) fits the data in comparison to the simple average (on the left graph), the closer the value of R^2 is to 1. The areas of the blue squares represent the squared residuals with respect to the linear regression. The areas of the red squares represent the squared residuals with respect to the average value.

- The sum of squares of residuals, also called the residual sum of squares:

$$SS_{\text{res}} = \sum_{i}(y_i - f_i)^2 = \sum_{i} e_i^2$$

- The total sum of squares (proportional to the variance of the data):

$$SS_{\text{tot}} = \sum_i (y_i - \bar{y})^2$$

The most general definition of the coefficient of determination is

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

In the best case, the modeled values exactly match the observed values, which results in $SS_{\text{res}} = 0$ and $R^2 = 1$. A baseline model, which always predicts $\bar{y}$, will have $R^2 = 0$. Models that have worse predictions than this baseline will have a negative R^2.

Relation to unexplained variance

In a general form, R^2 can be seen to be related to the fraction of variance unexplained (FVU), since the second term compares the unexplained variance (variance of the model's errors) with the total variance (of the data):

$$R^2 = 1 - \text{FVU}$$

As explained variance

Suppose R^2 = 0.49. This implies that 49% of the variability of the dependent variable in the data set has been accounted for, and the remaining 51% of the variability is still unaccounted for. For regression models, the regression sum of squares, also called the explained sum of squares, is defined as

$$SS_{\text{reg}} = \sum_i (f_i - \bar{y})^2$$

In some cases, as in simple linear regression, the total sum of squares equals the sum of the two other sums of squares defined above:

$$SS_{\text{res}} + SS_{\text{reg}} = SS_{\text{tot}}$$

See Partitioning in the general OLS model for a derivation of this result for one case where the relation holds. When this relation does hold, the above definition of R^2 is equivalent to

$$R^2 = \frac{SS_{\text{reg}}}{SS_{\text{tot}}} = \frac{SS_{\text{reg}}/n}{SS_{\text{tot}}/n}$$

where n is the number of observations (cases) on the variables.

In this form R^2 is expressed as the ratio of the explained variance (variance of the model's predictions, which is SS_{reg} / n) to the total variance (sample variance of the dependent variable, which is SS_{tot} / n).

This partition of the sum of squares holds for instance when the model values f_i have been obtained by linear regression. A milder sufficient condition reads as follows: The model has the form

$$f_i = \widehat{\alpha} + \widehat{\beta} q_i$$

where the q_i are arbitrary values that may or may not depend on i or on other free parameters (the common choice $q_i = x_i$ is just one special case), and the coefficient estimates $\widehat{\alpha}$ and $\widehat{\beta}$ are obtained by minimizing the residual sum of squares.

This set of conditions is an important one and it has a number of implications for the properties of the fitted residuals and the modelled values. In particular, under these conditions:

$$\bar{f} = \bar{y}.$$

As squared correlation coefficient

In linear least squares multiple regression with an estimated intercept term, R^2 equals the square of the Pearson correlation coefficient between the observed y and modeled (predicted) f data values of the dependent variable.

In a linear least squares regression with an intercept term and a single explanator, this is also equal to the squared Pearson correlation coefficient of the dependent variable y and explanatory variable x.

It should not be confused with the correlation coefficient between two estimates, defined as

$$\rho_{\widehat{\alpha},\widehat{\beta}} = \frac{\operatorname{cov}\left(\widehat{\alpha}, \widehat{\beta}\right)}{\sigma_{\widehat{\alpha}} \sigma_{\widehat{\beta}}},$$

where the covariance between two coefficient estimates, as well as their standard deviations, are obtained from the covariance matrix of the coefficient estimates.

Under more general modeling conditions, where the predicted values might be generated from a model different from linear least squares regression, an R^2 value can be calculated as the square of the correlation coefficient between the original y and modeled f data values. In this case, the value is not directly a measure of how good the modeled values are, but rather a measure of how good a predictor might be constructed from the modeled values (by creating a revised predictor of the form $\alpha + \beta f_i$). According to Everitt,[10] this usage is specifically the definition of the term "coefficient of determination": the square of the correlation between two (general) variables.

Interpretation

R^2 is a statistic that will give some information about the goodness of fit of a model. In regression, the R^2 coefficient of determination is a statistical measure of how well the regression predictions approximate the real data points. An R^2 of 1 indicates that the regression predictions perfectly fit the data.

Values of R^2 outside the range 0 to 1 can occur when the model fits the data worse than a horizontal hyperplane. This would occur when the wrong model was chosen, or nonsensical constraints were applied by mistake. If equation 1 of Kvålseth[11] is used (this is the equation used most often), R^2 can be less than zero. If equation 2 of Kvålseth is used, R^2 can be greater than one.

In all instances where R^2 is used, the predictors are calculated by ordinary least-squares regression: that is, by minimizing SS_{res}. In this case, R^2 increases as the number of variables in the model is increased (R^2 is monotone increasing with the number of variables included—it will never decrease). This illustrates a drawback to one possible use of R^2, where one might keep adding variables (Kitchen sink regression) to increase the R^2 value. For example, if one is trying to predict the sales of a model of car from the car's gas mileage, price, and engine power, one can include such irrelevant factors as the first letter of the model's name or the height of the lead engineer designing the car because the R^2 will never decrease as variables are added and will probably experience an increase due to chance alone.

This leads to the alternative approach of looking at the adjusted R^2. The explanation of this statistic is almost the same as R^2 but it penalizes the statistic as extra variables are included in the model. For cases other than fitting by ordinary least squares, the R^2 statistic can be calculated as above and may still be a useful measure. If fitting is by weighted least squares or generalized least squares, alternative versions of R^2 can be calculated appropriate to those statistical frameworks, while the "raw" R^2 may still be useful if it is more easily interpreted. Values for R^2 can be calculated for any type of predictive model, which need not have a statistical basis.

In a multiple linear model

Consider a linear model with more than a single explanatory variable, of the form

$$Y_i = \beta_0 + \sum_{j=1}^{p} \beta_j X_{i,j} + \varepsilon_i,$$

where, for the ith case, Y_i is the response variable, $X_{i,1}, \ldots, X_{i,p}$ are p regressors, and ε_i is a mean zero error term. The quantities $\beta_0, \ldots, \beta_p$ are unknown coefficients, whose values are estimated by least squares. The coefficient of determination R^2 is a measure of the global fit of the model. Specifically, R^2 is an element of [0, 1] and represents the proportion of variability in Y_i that may be attributed to some linear combination of the regressors (explanatory variables) in X.[12]

R^2 is often interpreted as the proportion of response variation "explained" by the regressors in the model. Thus, $R^2 = 1$ indicates that the fitted model explains all variability in y, while $R^2 = 0$ indicates no 'linear' relationship (for straight line regression, this means that the straight line model is a constant line (slope = 0, intercept = $\bar{y}$) between the response variable and regressors). An interior value such as $R^2 = 0.7$ may be interpreted as follows: "Seventy percent of the variance in the response variable can be explained by the explanatory variables. The remaining thirty percent can be attributed to unknown, lurking variables or inherent variability."

A caution that applies to R^2, as to other statistical descriptions of correlation and association is that "correlation does not imply causation." In other words, while correlations may sometimes provide valuable clues in uncovering causal relationships among variables, a non-zero estimated correlation between two variables is not, on its own, evidence that changing the value of one variable would result in changes in the values of other variables. For example, the practice of carrying matches (or a lighter) is correlated with incidence of lung cancer, but carrying matches does not cause cancer (in the standard sense of "cause").

In case of a single regressor, fitted by least squares, R^2 is the square of the Pearson product-moment correlation coefficient relating the regressor and the response variable. More generally, R^2 is the square of the correlation between the constructed predictor and the response variable. With more than one regressor, the R^2 can be referred to as the coefficient of multiple determination.

Inflation of R^2

In least squares regression using typical data, R^2 is at least weakly increasing with increases in the number of regressors in the model. Because increases in the number of regressors increase the value of R^2, R^2 alone cannot be used as a meaningful comparison of models with very different numbers of independent variables. For a meaningful comparison between two models, an F-test can be performed on the residual sum of squares, similar to the F-tests in Granger causality, though this is not always appropriate. As a reminder of this, some authors denote R^2 by R_q^2, where q is the number of columns in X (the number of explanators including the constant).

To demonstrate this property, first recall that the objective of least squares linear regression is

$$\min_b SS_{\text{res}}(b) \Rightarrow \min_b \sum_i (y_i - X_i b)^2$$

where X_i is a row vector of values of explanatory variables for case i and b is a column vector of coefficients of the respective elements of X_i.

The optimal value of the objective is weakly smaller as more explanatory variables are added and hence additional columns of X (the explanatory data matrix whose ith row is X_i) are added, by the fact that less constrained minimization leads to an optimal cost which is weakly smaller than more constrained minimization does. Given the previous conclusion and noting that SS_{tot} depends only on y, the non-decreasing property of R^2 follows directly from the definition above.

The intuitive reason that using an additional explanatory variable cannot lower the R^2 is this: Minimizing SS_{res} is equivalent to maximizing R^2. When the extra variable is included, the data always have the option of giving it an estimated coefficient of zero, leaving the predicted values and the R^2 unchanged. The only way that the optimization problem will give a non-zero coefficient is if doing so improves the R^2.

Caveats

R^2 does not indicate whether:

- the independent variables are a cause of the changes in the dependent variable;
- omitted-variable bias exists;
- the correct regression was used;
- the most appropriate set of independent variables has been chosen;
- there is collinearity present in the data on the explanatory variables;
- the model might be improved by using transformed versions of the existing set of independent variables;
- there are enough data points to make a solid conclusion.

Extensions

Adjusted R^2

The use of an adjusted R^2 (one common notation is $\bar{R}^2$, pronounced "R bar squared"; another is R_{adj}^2) is an attempt to account for the phenomenon of the R^2 automatically and spuriously increasing when extra explanatory variables are added to the model. There are many different

ways of adjusting (see [13] for an overview). By far the most used one, to the point that it is typically just referred to as adjusted R, is the correction proposed by Mordecai Ezekiel.[13][14] The adjusted R^2 (according to Ezekiel) is defined as

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1}$$

where p is the total number of explanatory variables in the model (not including the constant term), and n is the sample size. It can also be written as:

$$\bar{R}^2 = 1 - \frac{SS_{\text{res}}/\text{df}_e}{SS_{\text{tot}}/\text{df}_t}$$

where df_t is the degrees of freedom $n - 1$ of the estimate of the population variance of the dependent variable, and df_e is the degrees of freedom $n - p - 1$ of the estimate of the underlying population error variance.

The adjusted R^2 can be negative, and its value will always be less than or equal to that of R^2. Unlike R^2, the adjusted R^2 increases only when the increase in R^2 (due to the inclusion of a new explanatory variable) is more than one would expect to see by chance. If a set of explanatory variables with a predetermined hierarchy of importance are introduced into a regression one at a time, with the adjusted R^2 computed each time, the level at which adjusted R^2 reaches a maximum, and decreases afterward, would be the regression with the ideal combination of having the best fit without excess/unnecessary terms.

Adjusted R^2 can be interpreted as a less biased estimator of the population R^2, whereas the observed sample R^2 is a positively biased estimate of the population value.[15] Adjusted R^2 is more appropriate when evaluating model fit (the variance in the dependent variable accounted for by the independent variables) and in comparing alternative models in the feature selection stage of model building.[15]

The principle behind the adjusted R^2 statistic can be seen by rewriting the ordinary R^2 as

$$R^2 = 1 - \frac{VAR_{\text{res}}}{VAR_{\text{tot}}}$$

where $\mathbf{VAR_{\text{res}}} = SS_{\text{res}}/n$ and $\mathbf{VAR_{\text{tot}}} = SS_{\text{tot}}/n$ are the sample variances of the estimated residuals and the dependent variable respectively, which can be seen as biased estimates of the population variances of the errors and of the dependent variable. These estimates are replaced by statistically unbiased versions: $\mathbf{VAR_{\text{res}}} = SS_{\text{res}}/(n-p-1)$ and $\mathbf{VAR_{\text{tot}}} = SS_{\text{tot}}/(n-1)$.

Despite using unbiased estimators for the population variances of the error and the dependent variable, adjusted R^2 is not an unbiased estimator of the population R^2,[15] which results by using the population variances of the errors and the dependent variable instead of estimating them. Ingram Olkin and John W. Pratt derived the Minimum-variance unbiased estimator for the population R^2,[16] which is known as Olkin-Pratt estimator. Comparisons of different approaches for adjusting R^2 concluded that in most situations either an approximate version of the Olkin-Pratt estimator [15] or the exact Olkin-Pratt estimator [17] should be preferred over (Ezekiel) adjusted R^2.

Coefficient of partial determination

The coefficient of partial determination can be defined as the proportion of variation that cannot be explained in a reduced model, but can be explained by the predictors specified in a full(er) model.[18][19][20] This coefficient is used to provide insight into whether or not one or more additional predictors may be useful in a more fully specified regression model.

The calculation for the partial R^2 is relatively straightforward after estimating two models and generating the ANOVA tables for them. The calculation for the partial R^2 is

$$\frac{SS_{\text{res, reduced}} - SS_{\text{res, full}}}{SS_{\text{res, reduced}}},$$

which is analogous to the usual coefficient of determination:

$$\frac{SS_{\text{tot}} - SS_{\text{res}}}{SS_{\text{tot}}}.$$

Generalizing and decomposing R^2 [21]

As explained above, model selection heuristics such as the Adjusted R^2 criterion and the F-test examine whether the total R^2 sufficiently increases to determine if a new regressor should be added to the model. If a regressor is added to the model that is highly correlated with other regressors which have already been included, then the total R^2 will hardly increase, even if the new regressor is of relevance. As a result, the above-mentioned heuristics will ignore relevant regressors when cross-correlations are high.

Alternatively, one can decompose a generalized version of R^2 to quantify the relevance of deviating from a hypothesis.[21] As Hoornweg (2018) shows, several shrinkage estimators – such as Bayesian linear regression, ridge regression, and the (adaptive) lasso – make use of this decomposition of R^2 when they gradually shrink parameters from the unrestricted OLS solutions towards the hypothesized values. Let us first define the linear regression model as
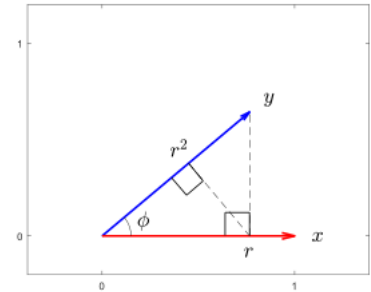
$$y = X\beta + \varepsilon.$$

Geometric representation of r^2.

It is assumed that the matrix X is standardized with Z-scores and that the column vector y is centered to have a mean of zero. Let the column vector β_0 refer to the hypothesized regression parameters and let the column vector b denote the estimated parameters. We can then define

$$R^2 = 1 - \frac{(y - Xb)'(y - Xb)}{(y - X\beta_0)'(y - X\beta_0)}.$$

An R^2 of 75% means that the in-sample accuracy improves by 75% if the data-optimized b solutions are used instead of the hypothesized β_0 values. In the special case that β_0 is a vector of zeros, we obtain the traditional R^2 again.

The individual effect on R^2 of deviating from a hypothesis can be computed with $R^{\otimes}$ ('R-outer'). This p times p matrix is given by

$$R^{\otimes} = (X'\tilde{y}_0)(X'\tilde{y}_0)'(X'X)^{-1}(\tilde{y}_0'\tilde{y}_0)^{-1},$$

where $\tilde{y}_0 = y - X\beta_0$. The diagonal elements of $R^{\otimes}$ exactly add up to R^2. If regressors are uncorrelated and β_0 is a vector of zeros, then the jth diagonal element of $R^{\otimes}$ simply corresponds to the r^2 value between x_j and y. When regressors x_i and x_j are correlated, $R_{ii}^{\otimes}$ might increase at the cost of a decrease in $R_{jj}^{\otimes}$. As a result, the diagonal elements of $R^{\otimes}$ may be smaller than 0 and, in more exceptional cases, larger than 1. To deal with such uncertainties, several shrinkage estimators implicitly take a weighted average of the diagonal elements of $R^{\otimes}$ to quantify the relevance of deviating from a hypothesized value.[21] Click on the lasso for an example.

R^2 in logistic regression

In the case of logistic regression, usually fit by maximum likelihood, there are several choices of pseudo-R^2.

One is the generalized R^2 originally proposed by Cox & Snell,[22] and independently by Magee:[23]

$$R^2 = 1 - \left(\frac{\mathcal{L}(0)}{\mathcal{L}(\hat{\theta})}\right)^{2/n}$$

where $\mathcal{L}(0)$ is the likelihood of the model with only the intercept, $\mathcal{L}(\hat{\theta})$ is the likelihood of the estimated model (i.e., the model with a given set of parameter estimates) and n is the sample size. It is easily rewritten to:

$$R^2 = 1 - e^{\frac{2}{n}(\ln(\mathcal{L}(0)) - \ln(\mathcal{L}(\hat{\theta})))} = 1 - e^{-D/n}$$

where D is the test statistic of the likelihood ratio test.

Nagelkerke[24] noted that it had the following properties:

1. It is consistent with the classical coefficient of determination when both can be computed;
2. Its value is maximised by the maximum likelihood estimation of a model;
3. It is asymptotically independent of the sample size;
4. The interpretation is the proportion of the variation explained by the model;
5. The values are between 0 and 1, with 0 denoting that model does not explain any variation and 1 denoting that it perfectly explains the observed variation;
6. It does not have any unit.

However, in the case of a logistic model, where $\mathcal{L}(\hat{\theta})$ cannot be greater than 1, R^2 is between 0 and $R_{\max}^2 = 1 - (\mathcal{L}(0))^{2/n}$: thus, Nagelkerke suggested the possibility to define a scaled R^2 as $R^2/R^2_{\max}$.[25]

Comparison with norm of residuals

Occasionally, the norm of residuals is used for indicating goodness of fit. This term is calculated as the square-root of the sum of squares of residuals:

$$\text{norm of residuals} = \sqrt{SS_{\text{res}}} = \|e\|.$$

Both R^2 and the norm of residuals have their relative merits. For least squares analysis R^2 varies between 0 and 1, with larger numbers indicating better fits and 1 representing a perfect fit. The norm of residuals varies from 0 to infinity with smaller numbers indicating better fits and zero indicating a perfect fit. One advantage and disadvantage of R^2 is the SS_{tot} term acts to normalize the value. If the y_i values are all multiplied by a constant, the norm of residuals will also change by that constant but R^2 will stay the same. As a basic example, for the linear least squares fit to the set of data:

$$
\begin{aligned}
x &= \quad 1, \quad 2, \quad 3, \quad 4, \quad 5 \\
y &= \quad 1.9, \quad 3.7, \quad 5.8, \quad 8.0, \quad 9.6
\end{aligned}
$$

R^2 = 0.998, and norm of residuals = 0.302. If all values of y are multiplied by 1000 (for example, in an SI prefix change), then R^2 remains the same, but norm of residuals = 302.

Another single-parameter indicator of fit is the RMSE of the residuals, or standard deviation of the residuals. This would have a value of 0.135 for the above example given that the fit was linear with an unforced intercept.[26]

History

The creation of the coefficient of determination has been attributed to the geneticist Sewall Wright and was first published in 1921.[27]

See also

- Anscombe's quartet
- Fraction of variance unexplained
- Goodness of fit
- Nash–Sutcliffe model efficiency coefficient (hydrological applications)
- Pearson product-moment correlation coefficient
- Proportional reduction in loss
- Regression model validation
- Root mean square deviation
- Stepwise regression
- t-test of $H_0 : R^2 = 0$.

Notes

1. Steel, R. G. D.; Torrie, J. H. (1960). *Principles and Procedures of Statistics with Special Reference to the Biological Sciences*. McGraw Hill.
2. Glantz, Stanton A.; Slinker, B. K. (1990). *Primer of Applied Regression and Analysis of Variance*. McGraw-Hill. ISBN 978-0-07-023407-9.
3. Draper, N. R.; Smith, H. (1998). *Applied Regression Analysis*. Wiley-Interscience. ISBN 978-0-471-17082-2.
4. Devore, Jay L. (2011). *Probability and Statistics for Engineering and the Sciences* (8th ed.). Boston, MA: Cengage Learning. pp. 508–510. ISBN 978-0-538-73352-6.
5. Barten, Anton P. (1987). "The Coeffecient of Determination for Regression without a Constant Term". In Heijmans, Risto; Neudecker, Heinz (eds.). *The Practice of Econometrics*. Dordrecht: Kluwer. pp. 181–189. ISBN 90-247-3502-5.

6. Colin Cameron, A.; Windmeijer, Frank A.G. (1997). "An R-squared measure of goodness of fit for some common nonlinear regression models". *Journal of Econometrics*. **77** (2): 1790–2. doi:10.1016/S0304-4076(96)01818-0 (https://doi.org/10.1016%2FS0304-4076%2896%2901818-0).

7. Chicco, Davide; Warrens, Matthijs J.; Jurman, Giuseppe (2021). "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation" (https://doi.org/10.7717%2Fpeerj-cs.623). *PeerJ Computer Science*. **7** (e623): 1–24. doi:10.7717/peerj-cs.623 (https://doi.org/10.7717%2Fpeerj-cs.623).

8. Legates, D.R.; McCabe, G.J. (1999). "Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation" (https://doi.org/10.1029%2F1998WR900018). *Water Resour. Res*. **35** (1): 233–241. Bibcode:1999WRR....35..233L (https://ui.adsabs.harvard.edu/abs/1999WRR....35..233L). doi:10.1029/1998WR900018 (https://doi.org/10.1029%2F1998WR900018).

9. Ritter, A.; Muñoz-Carpena, R. (2013). "Performance evaluation of hydrological models: statistical significance for reducing subjectivity in goodness-of-fit assessments". *Journal of Hydrology*. **480** (1): 33–45. Bibcode:2013JHyd..480...33R (https://ui.adsabs.harvard.edu/abs/2013JHyd..480...33R). doi:10.1016/j.jhydrol.2012.12.004 (https://doi.org/10.1016%2Fj.jhydrol.2012.12.004).

10. Everitt, B. S. (2002). *Cambridge Dictionary of Statistics* (2nd ed.). CUP. p. 78. ISBN 978-0-521-81099-9.

11. Kvalseth, Tarald O. (1985). "Cautionary Note about R2". *The American Statistician*. **39** (4): 279–285. doi:10.2307/2683704 (https://doi.org/10.2307%2F2683704). JSTOR 2683704 (https://www.jstor.org/stable/2683704).

12. Computing Adjusted R2 for Polynomial Regressions (https://www.mathworks.com/help/matlab/data_analysis/linear-regression.html#bswinlz)

13. Raju, Nambury S.; Bilgic, Reyhan; Edwards, Jack E.; Fleer, Paul F. (1997). "Methodology review: Estimation of population validity and cross-validity, and the use of equal weights in prediction" (https://doi.org/10.1177/01466216970214001). *Applied Psychological Measurement*. **21** (4): 291–305. doi:10.1177/01466216970214001 (https://doi.org/10.1177%2F01466216970214001). ISSN 0146-6216 (https://www.worldcat.org/issn/0146-6216). S2CID 122308344 (https://api.semanticscholar.org/CorpusID:122308344).

14. Yin, Ping; Fan, Xitao (January 2001). "Estimating R^2 Shrinkage in Multiple Regression: A Comparison of Different Analytical Methods" (http://www.tandfonline.com/doi/abs/10.1080/00220970109600656). *The Journal of Experimental Education*. **69** (2): 203–224. doi:10.1080/00220970109600656 (https://doi.org/10.1080%2F00220970109600656). ISSN 0022-0973 (https://www.worldcat.org/issn/0022-0973). S2CID 121614674 (https://api.semanticscholar.org/CorpusID:121614674). Retrieved 2021-04-23.

15. Shieh, Gwowen (2008-04-01). "Improved shrinkage estimation of squared multiple correlation coefficient and squared cross-validity coefficient". *Organizational Research Methods*. **11** (2): 387–407. doi:10.1177/1094428106292901 (https://doi.org/10.1177%2F1094428106292901). ISSN 1094-4281 (https://www.worldcat.org/issn/1094-4281). S2CID 55098407 (https://api.semanticscholar.org/CorpusID:55098407).

16. Olkin, Ingram; Pratt, John W. (March 1958). "Unbiased estimation of certain correlation coefficients" (https://projecteuclid.org/euclid.aoms/1177706717). *The Annals of Mathematical Statistics*. **29** (1): 201–211. doi:10.1214/aoms/1177706717 (https://doi.org/10.1214%2Faoms%2F1177706717). ISSN 0003-4851 (https://www.worldcat.org/issn/0003-4851).

17. Karch, Julian (2020-09-29). "Improving on Adjusted R-Squared" (https://doi.org/10.1525%2Fcollabra.343). *Collabra: Psychology*. **6** (45). doi:10.1525/collabra.343 (https://doi.org/10.1525%2Fcollabra.343). ISSN 2474-7394 (https://www.worldcat.org/issn/2474-7394).

18. Richard Anderson-Sprecher, "Model Comparisons and R^2 (http://www.tandfonline.com/doi/abs/10.1080/00031305.1994.10476036)", *The American Statistician*, Volume 48, Issue 2, 1994, pp. 113–117.

19. (generalized to Maximum Likelihood) N. J. D. Nagelkerke, "A Note on a General Definition of the Coefficient of Determination (http://www.cesarzamudio.com/uploads/1/7/9/1/17916581/nagelkerke_n.j.d._1991_-_a_note_on_a_general_definition_of_the_coefficient_of_determination.pdf)", *Biometrika*, Vol. 78, No. 3. (Sep., 1991), pp. 691–692.

20. "R implementation of coefficient of partial determination" (https://stats.stackexchange.com/q/7775)

21. Hoornweg, Victor (2018). "Part II: On Keeping Parameters Fixed" (http://www.victorhoornweg.com). *Science: Under Submission*. Hoornweg Press. ISBN 978-90-829188-0-9.

22. Cox, D. D.; Snell, E. J. (1989). *The Analysis of Binary Data* (2nd ed.). Chapman and Hall.

23. Magee, L. (1990). "R^2 measures based on Wald and likelihood ratio joint significance tests". *The American Statistician*. **44**. pp. 250–3. doi:10.1080/00031305.1990.10475731 (https://doi.org/10.1080%2F00031305.1990.10475731).

24. Nagelkerke, Nico J. D. (1992). *Maximum Likelihood Estimation of Functional Relationships, Pays-Bas*. Lecture Notes in Statistics. **69**. ISBN 978-0-387-97721-8.

25. Nagelkerke, N. J. D. (1991). "A Note on a General Definition of the Coefficient of Determination". *Biometrika*. **78** (3): 691–2. doi:10.1093/biomet/78.3.691 (https://doi.org/10.1093%2Fbiomet%2F78.3.691). JSTOR 2337038 (https://www.jstor.org/stable/2337038).

26. OriginLab webpage, http://www.originlab.com/doc/Origin-Help/LR-Algorithm. Retrieved February 9, 2016.

27. Wright, Sewall (January 1921). "Correlation and causation". *Journal of Agricultural Research*. **20**: 557–585.

Further reading

- Gujarati, Damodar N.; Porter, Dawn C. (2009). *Basic Econometrics* (Fifth ed.). New York: McGraw-Hill/Irwin. pp. 73–78. ISBN 978-0-07-337577-9.

- Hughes, Ann; Grawoig, Dennis (1971). *Statistics: A Foundation for Analysis* (https://archive.org/details/trent_0116302260611/page/344). Reading: Addison-Wesley. pp. 344–348 (https://archive.org/details/trent_0116302260611/page/344). ISBN 0-201-03021-7.

- Kmenta, Jan (1986). *Elements of Econometrics* (https://archive.org/details/elementsofeconom0003kmen/page/240) (Second ed.). New York: Macmillan. pp. 240–243 (https://archive.org/details/elementsofeconom0003kmen/page/240). ISBN 978-0-02-365070-3.

- Lewis-Beck, Michael S.; Skalaban, Andrew (1990). "The *R*-Squared: Some Straight Talk". *Political Analysis*. **2**: 153–171. doi:10.1093/pan/2.1.153 (https://doi.org/10.1093%2Fpan%2F2.1.153). JSTOR 23317769 (https://www.jstor.org/stable/23317769).

- Chicco, Davide; Warrens, Matthijs J.; Jurman, Giuseppe (2021). "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation" (https://doi.org/10.7717%2Fpeerj-cs.623). *PeerJ Computer Science*. **7** (e623): 1–24. doi:10.7717/peerj-cs.623 (https://doi.org/10.7717%2Fpeerj-cs.623).