Data Analytics

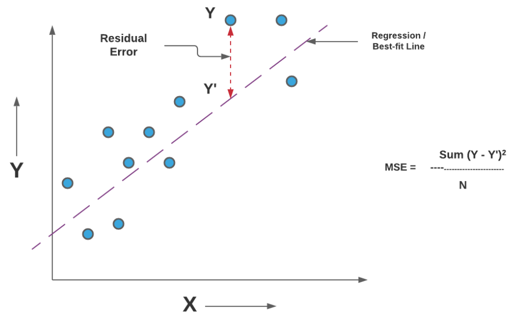
Data, Data Science, Machine Learning, AI

- Home
- Interview Questions
- Machine Learning »
- AI Companies

Select a page

Mean Squared Error or R-Squared – Which one to use?

September 30, 2020 by Ajitesh Kumar · Leave a comment



In this post, you will learn about the concepts of **mean-squared error (MSE)** and **R-squared**, difference between them and which one to use when working with regression models such as linear regression model. You also learn Python examples to understand the concepts in a better manner. In this post, the following topics are covered:

- Introduction to Mean Squared Error (MSE) and R-Squared
- Difference between MSE and R-Squared
- MSE or R-Squared – Which one to use?
- MSE and R-Squared Python code example

Table of Contents

Introduction to Mean Square Error (MSE) and R-Squared

In this section, you will learn about the concepts of mean squared error and R-squared. These are used for evaluating the performance of regression models such as linear regression model.

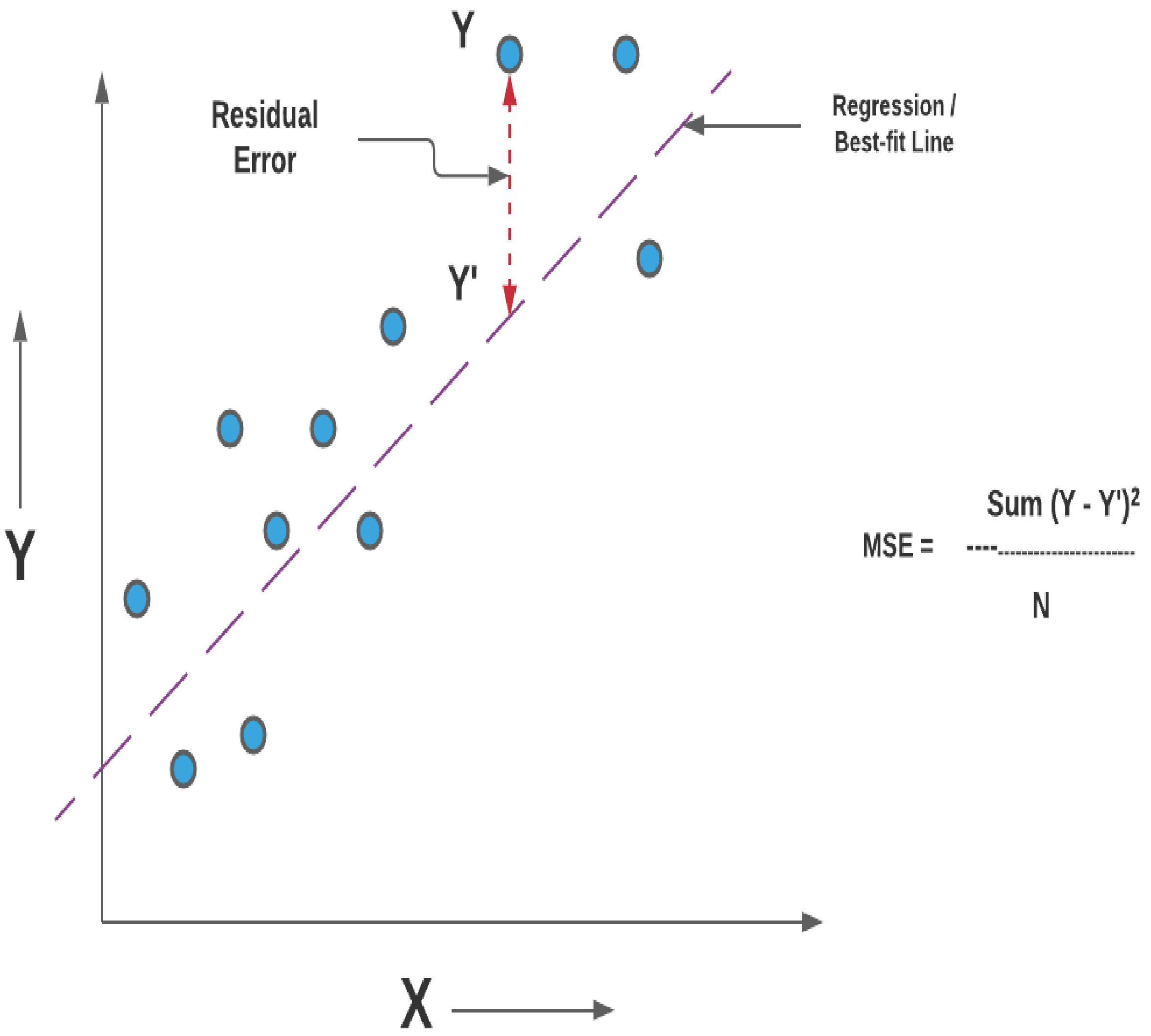
What is Mean Squared Error (MSE)?

Mean squared error (MSE) is the average of sum of squared difference between actual value and the predicted or estimated value. It is also termed as **mean squared deviation (MSD).** This is how it is represented mathematically:

$$MSE = \frac{1}{n}\sum\Big(\underbrace{y - \hat{y}}_{\text{The square of the difference between actual and predicted}}\Big)^2$$

Fig 1. Mean Squared Error

The value of MSE is always positive or greater than zero. A value close to zero will represent better quality of the estimator / predictor (regression model). **An MSE of zero (0) represents the fact that the predictor is a perfect predictor**. When you take a **square root of MSE** value, it becomes **root mean squared error (RMSE).** In the above equation, Y represents the actual value and the Y' is predicted value. Here is the diagrammatic representation of MSE:



Subscribe

Fig 2. Mean Squared Error Representation

What is R-Squared?

R-Squared is the ratio of Sum of Squares Regression (SSR) and Sum of Squares Total (SST). Sum of Squares Regression is amount of variance explained by the regression line. R-squared value is used to measure the **goodness of fit.** Greater the value of R-Squared, better is the regression model. However, we need to take a caution. This is where **adjusted R-squared** concept comes into picture. This would be discussed in one of the later posts. R-Squared is also termed as the **coefficient of determination.** For the training dataset, the is bounded between 0 and 1, but it can become negative for the test dataset if the SSE is greater than SST. If the value of R-Squared is 1, the model fits the data perfectly with a corresponding $MSE = 0$.

Here is a visual representation to understand the concepts of R-Squared in a better manner.

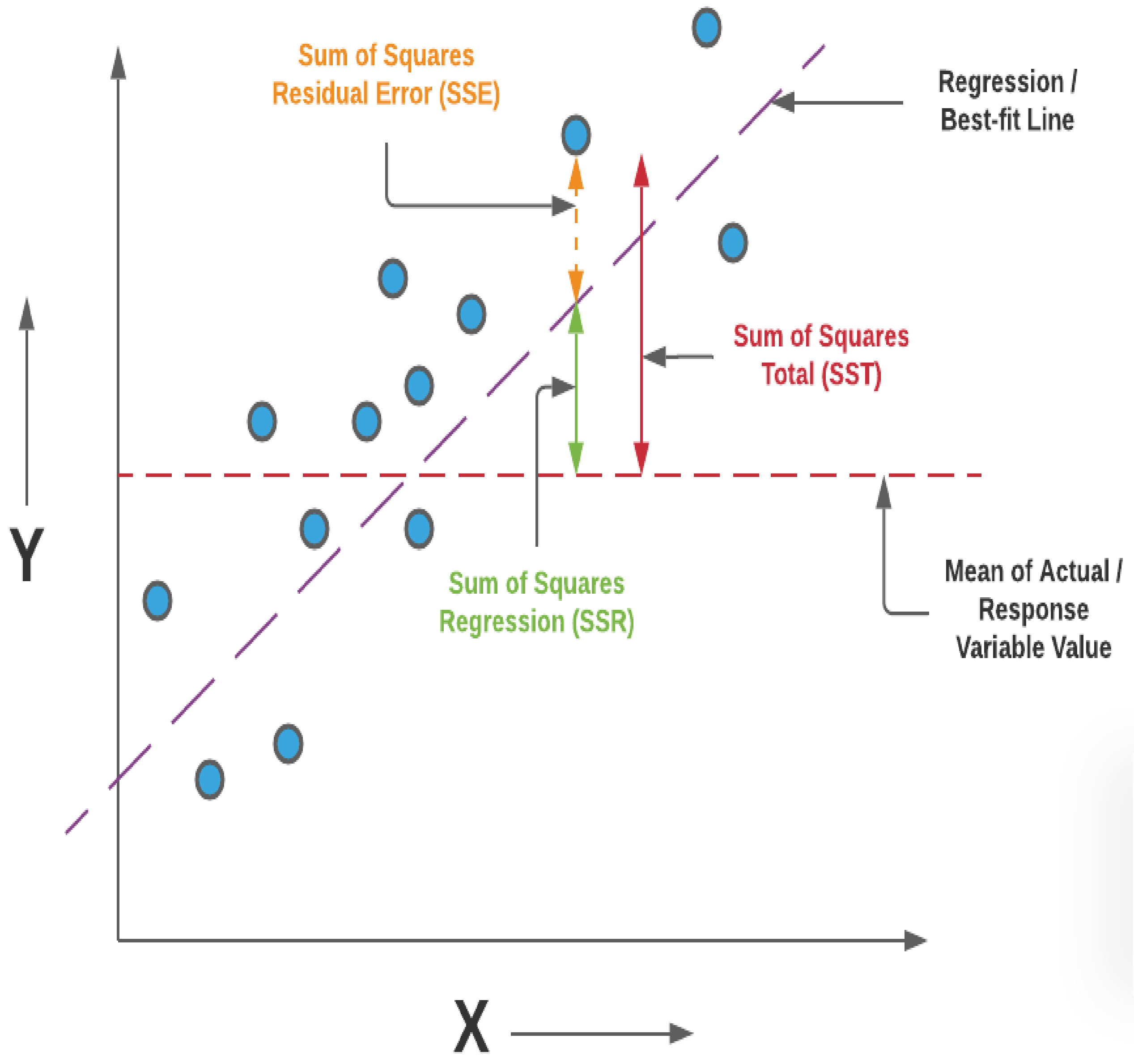


Fig 4. Diagrammatic representation for understanding R-Squared

$$R^2 = \frac{SSR}{SST} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$$

Pay attention to the diagram and note that greater the value of SSR, more is the variance covered by the regression / best fit line out of total variance (SST). R-Squared can also be represented using the following formula:

R-Squared = 1 – (SSE/SST)

Pay attention to the diagram and note that smaller the value of SSE, smaller is the value of (SSE/SST) and hence greater will be value of R-Squared.

R-Squared can also be expressed as a function of mean squared error (MSE). The following equation represents the same.

$$\begin{aligned} R^2 &= 1 - \frac{SSE}{SST} \\ &= \frac{\frac{1}{n}\sum_{i=1}^{n}\left(y^{(i)} - \hat{y}^{(i)}\right)^2}{\frac{1}{n}\sum_{i=1}^{n}\left(y^{(i)} - \mu_y\right)^2} \\ &= 1 - \frac{MSE}{Var(y)} \end{aligned}$$

Difference between Mean Square Error & R-Squared

The similarity between mean-squared error and R-Squared is that they both are a type of metrics which are used for evaluating the performance of the regression models, especially statistical model such as linear regression model. The difference is that MSE gets pronounced based on whether the data is scaled or not. For example, if the response variable is housing price in the multiple of 10K, MSE will be different (lower) than when the response variable such as housing pricing is not scaled (actual values). This is where R-Squared comes to the rescue. R-Squared is also termed as the standardized version of MSE. R-squared represents the fraction of variance of response variable captured by the regression model ratł than the MSE which captures the residual error.

MSE or R-Squared – Which one to Use?

It is recommended to use R-Squared or rather adjusted R-Squared for evaluating the model performance of the regression models. This is primarily because R-Squared captures the fraction of response variance captured by the regression and tend to give better picture of quality of regression model. Also, MSE values differ based on whether the values of the response variable is scaled or not. A better measure is root mean squared error (RMSE) which takes care of the fact related to whether the values of the response variable is scaled or not.

One can alternatively use MSE or R-Squared based on what is appropriate and need of the hour.

MSE or R-Squared Python Code Example

Here is the python code representing how to calculate mean squared error or R-Squared value while working with regression models. Pay attention to some of the following in the code given below:

- **Sklearn.metrics mean_squared_error** and **r2_score** is used for measuring the MSE and R-Squared values. Input to this methods are actual values and predicted values.
- Sklearn Boston housing dataset is used for training a multiple linear regression model using Sklearn.linear_model LinearRegression

```
import pandas as pd
from sklearn.model_selection import train_test_split
```

```
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LinearRegression
from sklearn.pipeline import make_pipeline
from sklearn.metrics import mean_squared_error, r2_score
from sklearn import datasets
#
# Load the Sklearn Boston Dataset
#
boston_ds = datasets.load_boston()
X = boston_ds.data
y = boston_ds.target
#
# Create a training and test split
#
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, rando
#
# Fit a pipeline using Training dataset and related labels
#
pipeline = make_pipeline(StandardScaler(), LinearRegression())
pipeline.fit(X_train, y_train)
#
# Calculate the predicted value for training and test dataset
#
y_train_pred = pipeline.predict(X_train)
y_test_pred = pipeline.predict(X_test)
#
# Mean Squared Error
#
print('MSE train: %.3f, test: %.3f' % (mean_squared_error(y_train, y_train_
                mean_squared_error(y_test, y_test_pred)))
#
# R-Squared
#
print('R^2 train: %.3f, test: %.3f' % (r2_score(y_train, y_train_pred),
                r2_score(y_test, y_test_pred)))
```


Conclusions

Here is the summary of what you learned in this post regarding **mean square error (MSE) and R-Squared** and which one to use?

- MSE represents the residual error which is nothing but sum of squared difference between actual values and the predicted / estimated values.
- R-Squared represents the fraction of response variance captured by the regression model
- The disadvantage of using MSE is that the value of MSE varies based on whether the values of response variable is scaled or not. If scaled, MSE will be lower than the unscaled values.

Author Recent Posts

Ajitesh Kumar

Follow me

I have been recently working in the area of Data Science and Machine Learning / Deep Learning. In addition, I am also passionate about various different technologies including programming languages such as Java/JEE, Javascript, Python, R, Julia etc and technologies such as Blockchain, mobile computing, cloud-native technologies, application security, cloud computing platforms, big data etc. I would love to connect with you on Linkedin.

Posted in Data Science, Machine Learning, Python. Tagged with Data Science, machine learning, python.

← Linear Regression Explained with Python Examples
RANSAC Regression Explained with Python Examples →

Leave a Reply

Your email address will not be published. Required fields are marked *

Comment

Name *

Email *

Website

Time limit is exhausted. Please reload the CAPTCHA.

eight + =

Post Comment

- Search

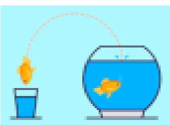
Subscribe

-
- Recent Posts
    - Machine Learning: Inference & Prediction Difference
    - What are Actionable Insights: Examples & Concepts
    - How to Create Data-Driven Culture: Key Steps
    - Overfitting & Underfitting Concepts & Interview Questions
    - Data Readiness Levels Assessment: Concepts

-
- Tags

  ai (66) Angular (50) angularjs (104) api (16) Application Security (22) artificial intelligence (20) AWS (23) bigdata (11) big data (40) blockchain (58) cloud (11) cloud computing (11) data analytics (19) Data Science (315) datascience (34) Deep Learning (34) docker (26) freshers (14) google (14) google glass (11) hyperledger (18) Interview questions (71) Java (92) javascript (103) Kubernetes (19) machine learning (309) mongodb (16) news (13) nlp (13) nosql (17) python (88) QA (12) quantum computing (13) reactjs (15) r programming (11) sklearn (30) Software Quality (11) spring framework (16) statistics (15) testing (16) tools (11) tutorials (14) UI (13) Unit Testing (18) web (16)

About Us



Vitalflux.com is dedicated to help software engineers & data scientists get technology news, practice tests, tutorials in order to reskill / acquire newer skills from time-to-time.

Thank you for visiting our site today. We welcome all your suggestions in order to make our website better. Please feel free to share your thoughts.

Latest Technologies

Machine Learning / Data Science
Quantum Computing
Robotics
Blockchain
Cloud Computing
Web

Subscribe

Statistics. Simplified.
STATOLOGY

MENU

FEBRUARY 22, 2021 BY ZACH

A Gentle Guide to Sum of Squares: SST, SSR, SSE

Linear regression is used to find a line that best "fits" a dataset.

We often use three different **sum of squares** values to measure how well the regression line actually fits the data:

1. Sum of Squares Total (SST) – The sum of squared differences between individual data points (y_i) and the mean of the response variable ($\overline{y}$).

- SST = $\Sigma(y_i - \overline{y})^2$

2. Sum of Squares Regression (SSR) – The sum of squared differences between predicted data points ($\hat{y}_i$) and the mean of the response variable($\overline{y}$).

- SSR = $\Sigma(\hat{y}_i - \overline{y})^2$

3. Sum of Squares Error (SSE) – The sum of squared differences between predicted data points ($\hat{y}_i$) and observed data points (y_i).

- SSE = $\Sigma(\hat{y}_i - y_i)^2$

The following relationship exists between these three measures:

SST = SSR + SSE

Thus, if we know two of these measures then we can use some simple algebra to calculate the third.

SSR, SST & R-Squared

R-squared, sometimes referred to as the coefficient of determination, is a measure of how well a linear regression model fits a dataset. It represents the proportion of the variance in the response variable that can be explained by the predictor variable.

The value for R-squared can range from 0 to 1. A value of 0 indicates that the response variable cannot be explained by the predictor variable at all. A value of 1 indicates that the response variable can be perfectly explained without error by the predictor variable.

Using SSR and SST, we can calculate R-squared as:

R-squared = SSR / SST

For example, if the SSR for a given regression model is 137.5 and SST is 156 then we would calculate R-squared as:

R-squared = 137.5 / 156 = 0.8814

This tells us that 88.14% of the variation in the response variable can be explained by the predictor variable.

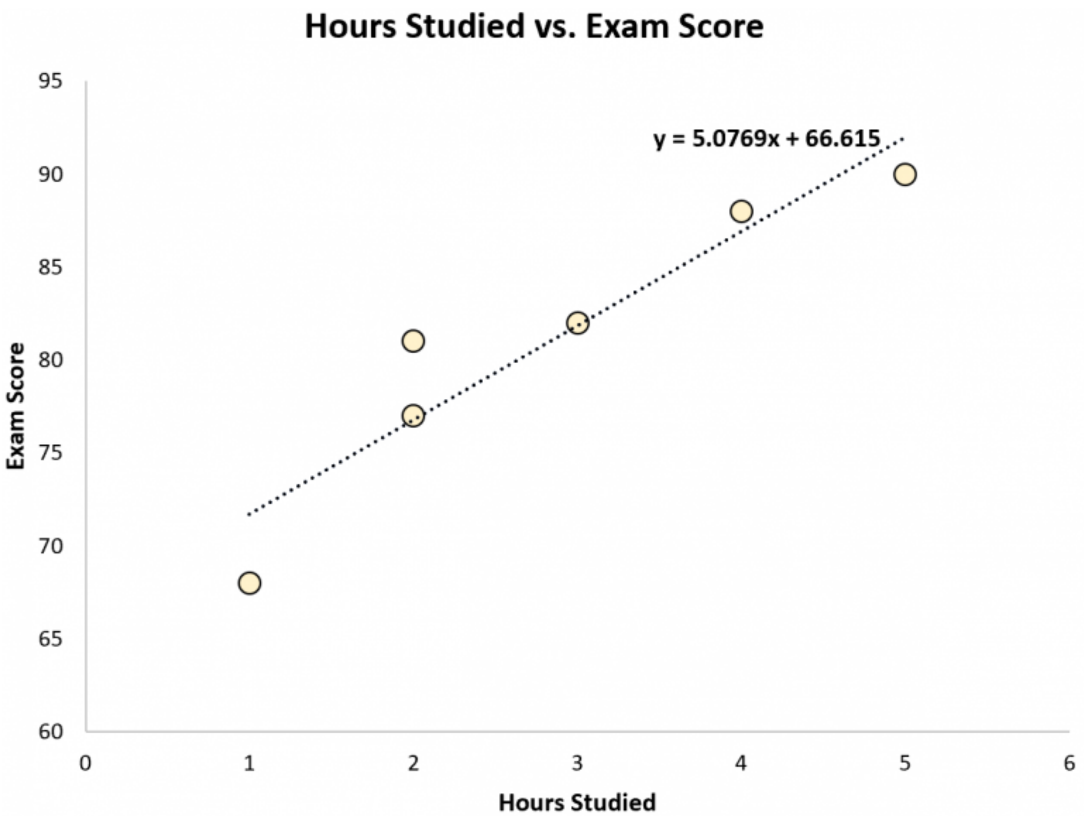
Calculate SST, SSR, SSE: Step-by-Step Example

Suppose we have the following dataset that shows the number of hours studied by six different students along with their final exam scores:

Hours Studied	Exam Score
1	68
2	77
2	81
3	82
4	88
5	90

Using some statistical software (like R, Excel, Python) or even by hand, we can find that the line of best fit is:

Score = 66.615 + 5.0769*(Hours)



Once we know the line of best fit equation, we can use the following steps to calculate SST, SSR, and SSE:

Step 1: Calculate the mean of the response variable.

The mean of the response variable ($\bar{y}$) turns out to be **81**.

Hours Studied	Exam Score	$\bar{y}$
1	68	81
2	77	81
2	81	81
3	82	81
4	88	81
5	90	81

Step 2: Calculate the predicted value for each observation.

Next, we can use the line of best fit equation to calculate the predicted exam score () for each student.

For example, the predicted exam score for the student who studied one hours is:

Score = 66.615 + 5.0769*(1) = **71.69**.

We can use the same approach to find the predicted score for each student:

Hours Studied	Exam Score	$\bar{y}$	$\hat{y}$
1	68	81	71.69
2	77	81	76.77
2	81	81	76.77
3	82	81	81.85
4	88	81	86.92
5	90	81	92.00

Step 3: Calculate the sum of squares total (SST).

Next, we can calculate the sum of squares total.

For example, the sum of squares total for the first student is:

$(y_i - \bar{y})^2 = (68 - 81)^2 =$ **169**.

We can use the same approach to find the sum of squares total for each student:

Hours Studied	Exam Score	$\bar{y}$	$\hat{y}$	$(y_i - \bar{y})^2$
1	68	81	71.69	169
2	77	81	76.77	16
2	81	81	76.77	0
3	82	81	81.85	1
4	88	81	86.92	49
5	90	81	92.00	81
				316
				SST

The sum of squares total turns out to be **316**.

Step 4: Calculate the sum of squares regression (SSR).

Next, we can calculate the sum of squares regression.

For example, the sum of squares regression for the first student is:

$(\hat{y}_i - \bar{y})^2 = (71.69 - 81)^2 =$ **86.64.**

We can use the same approach to find the sum of squares regression for each student:

Hours Studied	Exam Score	$\bar{y}$	$\hat{y}$	$(y_i - \bar{y})^2$	$(\hat{y}_i - \bar{y})^2$
1	68	81	71.69	169	86.64
2	77	81	76.77	16	17.90
2	81	81	76.77	0	17.90
3	82	81	81.85	1	0.72
4	88	81	86.92	49	35.08
5	90	81	92.00	81	120.99
				316	279.23
				SST	**SSR**

The sum of squares regression turns out to be **279.23**.

Step 5: Calculate the sum of squares error (SSE).

Next, we can calculate the sum of squares error.

For example, the sum of squares error for the first student is:

$(\hat{y}_i - y_i)^2 = (71.69 - 68)^2 = \mathbf{13.63}$.

We can use the same approach to find the sum of squares error for each student:

Hours Studied	Exam Score	$\bar{y}$	$\hat{y}$	$(y_i - \bar{y})^2$	$(\hat{y}_i - \bar{y})^2$	$(\hat{y}_i - y_i)^2$
1	68	81	71.69	169	86.64	13.63
2	77	81	76.77	16	17.90	0.05
2	81	81	76.77	0	17.90	17.90
3	82	81	81.85	1	0.72	0.02
4	88	81	86.92	49	35.08	1.16
5	90	81	92.00	81	120.99	4.00
				316	279.23	36.77
				SST	**SSR**	**SSE**

We can verify that SST = SSR + SSE

- SST = SSR + SSE
- 316 = 279.23 + 36.77

We can also calculate the R-squared of the regression model by using the following equation:

- R-squared = SSR / SST
- R-squared = 279.23 / 316
- R-squared = 0.8836

This tells us that **88.36%** of the variation in exam scores can be explained by the number of hours studied.

Additional Resources

You can use the following calculators to automatically calculate SST, SSR, and SSE for any simple linear regression line:

SST Calculator
SSR Calculator
SSE Calculator

Published by Zach

View all posts by Zach

PREV

What Are Disjoint Events? (Definition & Examples)

NEXT

How to Calculate SST, SSR, and SSE in Excel

Leave a Reply

Your email address will not be published. Required fields are marked *

Comment

Name *

Email *

Website

POST COMMENT

SEARCH

Search ...

ABOUT

Statology is a site that makes learning statistics easy by explaining topics in simple and straightforward ways. **Learn more about us**.

STATOLOGY STUDY

Statology Study is the ultimate online statistics study guide that helps you understand all of the core concepts taught in any elementary statistics course and makes your life so much easier as a student.

STATOLOGY STUDY

CALCULATORS

Try out our free **online statistics calculators** if you're looking for some help finding probabilities, p-values, critical values, sample sizes, expected values, summary statistics, or correlation coefficients.

RECENT POSTS

Calculating the Difference Between Two Dates in Google Sheets

How to Use the PMT Function in Google Sheets (3 Examples)

How to Calculate Compound Interest in Google Sheets (3 Examples)

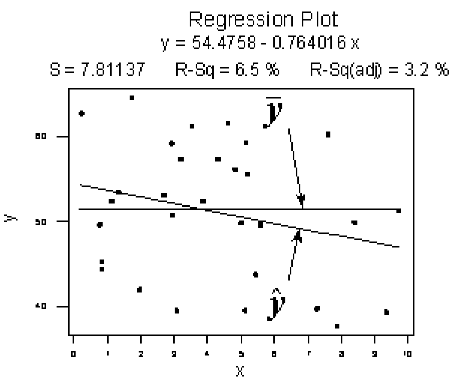
STAT 462

Applied Regression Analysis

2.5 - The Coefficient of Determination, r-squared

Let's start our investigation of the coefficient of determination, r^2, by looking at two different examples — one example in which the relationship between the response y and the predictor x is very weak and a second example in which the relationship between the response y and the predictor x is fairly strong. If our measure is going to work well, it should be able to distinguish between these two very different situations.

Here's a plot illustrating a very weak relationship between y and x. There are two lines on the plot, a horizontal line placed at the average response, $\bar{y}$, and a shallow-sloped estimated regression line, $\hat{y}$. Note that the slope of the estimated regression line is not very steep, suggesting that as the predictor x increases, there is not much of a change in the average response y. Also, note that the data points do not "hug" the estimated regression line:



$$SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = 119.1$$

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 1708.5$$

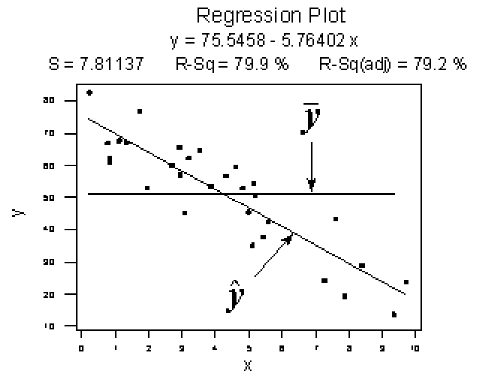
$$SSTO = \sum_{i=1}^{n}(y_i - \bar{y})^2 = 1827.6$$

The calculations on the right of the plot show contrasting "sums of squares" values:

- SSR is the "regression sum of squares" and quantifies how far the estimated sloped regression line, $\hat{y}_i$, is from the horizontal "no relationship line," the sample mean or $\bar{y}$.
- SSE is the "error sum of squares" and quantifies how much the data points, y_i, vary around the estimated regression line, $\hat{y}_i$.
- SSTO is the "total sum of squares" and quantifies how much the data points, y_i, vary around their mean, $\bar{y}$.

Note that SSTO = SSR + SSE. The sums of squares appear to tell the story pretty well. They tell us that most of the variation in the response y (*SSTO* = 1827.6) is just due to random variation (*SSE* = 1708.5), not due to the regression of y on x (*SSR* = 119.1). You might notice that *SSR* divided by *SSTO* is 119.1/1827.6 or 0.065. Do you see where this quantity appears on the above fitted line plot?

Contrast the above example with the following one in which the plot illustrates a fairly convincing relationship between y and x. The slope of the estimated regression line is much steeper, suggesting that as the predictor x increases, there is a fairly substantial change (decrease) in the response y. And, here, the data points do "hug" the estimated regression line:



$$SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = 6679.3$$

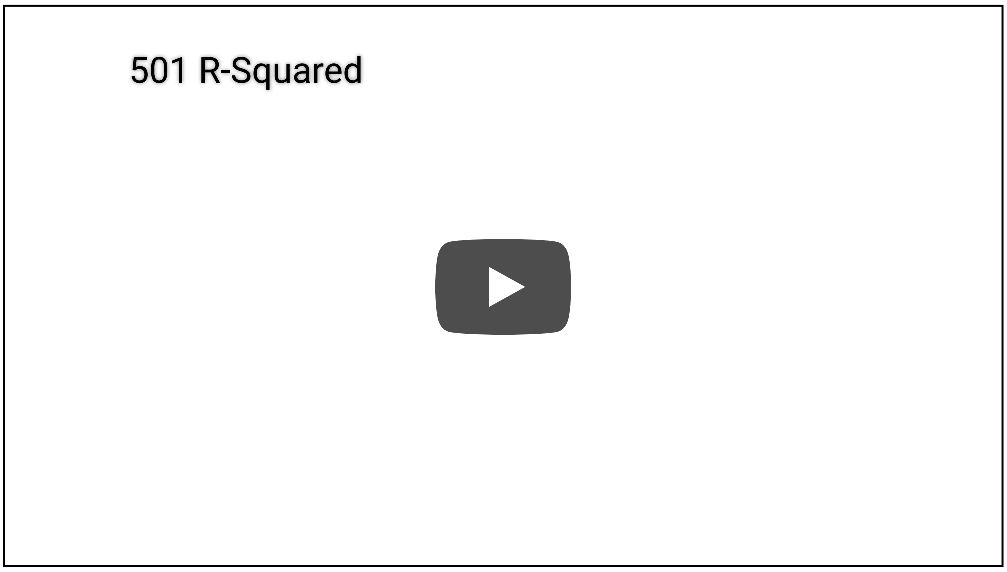
$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 1708.5$$

$$SSTO = \sum_{i=1}^{n}(y_i - \bar{y})^2 = 8487.8$$

The sums of squares for this dataset tell a very different story, namely that most of the variation in the response y ($SSTO$ = 8487.8) is due to the regression of y on x (SSR = 6679.3) not just due to random error (SSE = 1708.5). And, SSR divided by $SSTO$ is 6679.3/8487.8 or 0.799, which again appears on the fitted line plot.

The previous two examples have suggested how we should define the measure formally. In short, the "**coefficient of determination**" or "***r*-squared value**," denoted r^2, is the regression sum of squares divided by the total sum of squares. Alternatively, as demonstrated in this screencast below, since $SSTO = SSR + SSE$, the quantity r^2 also equals one minus the ratio of the error sum of squares to the total sum of squares:

$$r^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}$$



Here are some basic characteristics of the measure:

- Since r^2 is a proportion, it is always a number between 0 and 1.
- If r^2 = 1, all of the data points fall perfectly on the regression line. The predictor x accounts for *all* of the variation in y!
- If r^2 = 0, the estimated regression line is perfectly horizontal. The predictor x accounts for *none* of the variation in y!

We've learned the interpretation for the two easy cases — when r^2 = 0 or r^2 = 1 — but, how do we interpret r^2 when it is some number between 0 and 1, like 0.23 or 0.57, say? Here are two similar, yet slightly different, ways in which the coefficient of determination r^2 can be interpreted. We say either:

"$r^2 \times 100$ percent of the variation in y is reduced by taking into account predictor x"

or:

"$r^2 \times 100$ percent of the variation in y is "explained by" the variation in predictor x."

Many statisticians prefer the first interpretation. I tend to favor the second. The risk with using the second interpretation — and hence why "explained by" appears in quotes — is that it can be misunderstood as suggesting that the predictor x *causes* the change in the response y. Association is not causation. That is, just because a dataset is characterized by having a large r-squared value, it does not imply that x *causes* the changes in y. As long as you keep the correct meaning in mind, it is fine to use the second interpretation. A variation on the second interpretation is to say, "$r^2 \times 100$ percent of the variation in y is accounted for by the variation in predictor x."

Students often ask: "what's considered a large r-squared value?" It depends on the research area. Social scientists who are often trying to learn something about the huge variation in human behavior will tend to find it very hard to get r-squared values much above, say 25% or 30%. Engineers, on the other hand, who tend to study more exact systems would likely find an r-squared value of just 30% unacceptable. The moral of the story is to read the literature to learn what typical r-squared values are for your research area!

Let's revisit the skin cancer mortality example (skincancer.txt (../../sites/onlinecourses.science.psu.edu.stat462/files/data/skincancer/index.txt)). Any statistical software that performs simple linear regression analysis will report the r-squared value for you, which in this case is 67.98% or 68% to the nearest whole number.

We can say that 68% of the variation in the skin cancer mortality rate is reduced by taking into account latitude. Or, we can say — with knowledge of what it really means — that 68% of the variation in skin cancer mortality is "explained by" latitude.

‹ 2.4 - What is the Common Error Variance? (../94/) up (../79/) 2.6 - (Pearson) Correlation Coefficient r › (../96/)

MAE, MSE, RMSE, Coefficient of Determination, Adjusted R Squared — Which Metric is Better?

A

Akshita Chugh Follow

Dec 8, 2020 · 4 min read

The objective of Linear Regression is to find a line that minimizes the prediction error of all the data points.

The essential step in any machine learning model is to evaluate the accuracy of the model. The Mean Squared Error, Mean absolute error, Root Mean Squared Error, and R-Squared or Coefficient of determination metrics are used to evaluate the performance of the model in regression analysis.

- The Mean absolute error represents the average of the absolute difference between the actual and predicted values in the dataset. It measures the average of the residuals in the dataset.

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}|$$

Where,

$\hat{y}$ — predicted value of y

$\bar{y}$ — mean value of y

- Mean Squared Error represents the average of the squared difference between the original and predicted values in the data set. It measures the variance of the residuals.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2$$

- Root Mean Squared Error is the square root of Mean Squared error. It measures the standard deviation of residuals.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2}$$

- The coefficient of determination or R-squared represents the proportion of the variance in the dependent variable which is explained by the linear regression model. It is a scale-free score i.e. irrespective of the values being small or large, the value of R square will be less than one.

$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2}$$

- Adjusted R squared is a modified version of R square, and it is adjusted for the number of independent variables in the model, and it will always be less than or equal to R^2.In the formula below **n** is the number of observations in the data and **k** is the number of the independent variables in the data.

$$R^2_{adj} = 1 - \left[\frac{(1-R^2)(n-1)}{n-k-1}\right]$$

Differences among these evaluation metrics

- Mean Squared Error(MSE) and Root Mean Square Error penalizes the large prediction errors vi-a-vis Mean Absolute Error (MAE). However, RMSE is widely used than MSE to evaluate the performance of the regression model with other random models as it has the same units as the dependent variable (Y-axis).

- MSE is a differentiable function that makes it easy to perform mathematical operations in comparison to a non-differentiable function like MAE. Therefore, in many models, RMSE is used as a default metric for calculating Loss Function despite being harder to interpret than MAE.

- MAE is more robust to data with outliers.

- The lower value of MAE, MSE, and RMSE implies higher accuracy of a regression model. However, a higher value of R square is considered desirable.

- R Squared & Adjusted R Squared are used for explaining how well the independent variables in the linear regression model explains the variability in the dependent variable. R Squared value always increases with the addition of the independent variables which might lead to the addition of the redundant variables in our model. However, the adjusted R-squared solves this problem.

- Adjusted R squared takes into account the number of predictor variables, and it is used to determine the number of independent variables in our model. The value of Adjusted R squared decreases if the increase in the R square by the additional variable isn't significant enough.

- For comparing the accuracy among different linear regression models, RMSE is a better choice than R Squared.

Conclusion

Therefore, if comparing the prediction accuracy among different linear regression (LR)models then RMSE is a better option as it is simple to calculate and differentiable. However, if your dataset has outliers then choose MAE over RMSE.

Besides, the number of predictor variables in a linear regression model is determined by adjusted R squared, and choose RMSE over adjusted R squared if you care about evaluating prediction accuracy among different LR models.

Linear Regression Machine Learning Data Science R Squared Root Mean Squared Error