

How can I calculate the Matthews Correlation Coefficient in R?”

Authored by
stats writer

June 26, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I calculate the Matthews Correlation Coefficient in R?”*.
PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=154311>

The Matthews Correlation Coefficient (MCC) is a measure of the quality of binary classifications that takes into account true positives, true negatives, false positives, and false negatives. In order to calculate the MCC in R, you can use the "mcc" function from the "caret" package. This function takes in the predicted values and the actual values as inputs and returns the MCC score. It is a simple yet effective way to assess the accuracy of a binary classification model in R. By using the "mcc" function, you can easily evaluate the performance of your model and make informed decisions based on the results.

Calculate Matthews Correlation Coefficient in R

Matthews correlation coefficient (MCC) is a metric we can use to assess the performance of a .

It is calculated as:

$$\text{MCC} = \frac{(\text{TP} * \text{TN}) - (\text{FP} * \text{FN})}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}$$

where:

TP: Number of true positives
TN: Number of true negatives
FP: Number of false positives
FN: Number of false negatives

This metric is particularly useful when the two classes are imbalanced - that is, one class appears much more than the other.

The value for MCC ranges from -1 to 1 where:

-1 indicates total disagreement between predicted classes and actual classes
 0 is synonymous with completely random guessing
 1 indicates total agreement between predicted classes and actual classes

For example, suppose a sports analyst uses a to predict whether or not 400 different college basketball players get drafted into the NBA.

The following confusion matrix summarizes the predictions made by the model:

		Predicted	
		Drafted = Yes	Drafted = No
Actual	Drafted = Yes	15 (True Positive)	5 (False Negative)
	Drafted = No	5 (False positive)	375 (True Negative)

To calculate the MCC of the model, we can use the following formula:

$$\text{MCC} = \frac{(\text{TP} * \text{TN}) - (\text{FP} * \text{FN})}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}$$

$$\text{MCC} = \frac{(15 * 375 - 5 * 5)}{\sqrt{(15 + 5)(15 + 5)(375 + 5)(375 + 5)}} = 0.7368$$

Matthews correlation coefficient turns out to be 0.7368.

This value is somewhat close to one, which indicates that the model does a decent job of predicting whether or not players will get drafted.

The following example shows how to calculate MCC for this exact scenario using the `mcc()` function from the `mltools` package in R.

Example: Calculating Matthews Correlation Coefficient in R

```
library(mltools)

#define vector of actual classes
actual <- rep(c(1, 0), times=c(20, 380))

#define vector of predicted classes
preds <- rep(c(1, 0, 1, 0), times=c(15, 5, 5, 375))

#calculate Matthews correlation coefficient
mcc(preds, actual)

0.7368421
```

Matthews correlation coefficient is 0.7368.

This matches the value that we calculated earlier by

hand.

If you'd like to calculate Matthews correlation coefficient for a confusion matrix, you can use the confusionM argument as follows:

```
library(mltools)
```

```
#create confusion matrix
```

```
conf_matrix <- matrix(c(15, 5, 5, 375), nrow=2)
```

```
#view confusion matrix
```

```
conf_matrix
```

```
15 5
```

```
5 375
```

```
#calculate Matthews correlation coefficient for  
confusion matrix
```

```
mcc(confusionM = conf_matrix)
```

```
0.7368421
```

Once again, Matthews correlation coefficient is 0.7368