

How to Choose Between MAE and RMSE for Accurate Predictions

Authored by
stats writer

December 3, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Choose Between MAE and RMSE for Accurate Predictions*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=104335>

When developing models aimed at predicting future outcomes or estimating unknown quantities, accurately assessing model performance is paramount. Two fundamental metrics dominate the discussion around measuring predictive accuracy in regression models: the Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE). While both statistics quantify the typical magnitude of prediction errors, their underlying mathematical properties lead to distinct interpretations and application scenarios. Understanding the nuances between MAE and RMSE is essential for data scientists, ensuring that the chosen metric aligns with the specific goals and tolerance for error in the practical application domain.

The choice between these two metrics often boils down to how one wishes to penalize large prediction mistakes. MAE, calculated as the average of the absolute differences between predictions and actual values, provides a linear score, meaning the penalty increases proportionally with the error magnitude. This makes it an excellent choice for evaluating forecasting accuracy when dealing with data that may contain significant outliers, as it is inherently less sensitive to extreme values. Its result is easily interpretable, representing the average error in the units of the response variable.

Conversely, RMSE, which involves squaring the residuals before averaging and taking the square root, imposes a non-linear penalty. Because large errors are squared, they are disproportionately emphasized in the final score. Consequently, RMSE is the preferred metric when even small errors are deemed important, or when the cost associated with large errors is significantly higher than the cost of small errors. The inherent difference in how these metrics handle residuals--absolute values versus squaring--is the key distinction dictating which metric is best suited for performance assessment in a given predictive modeling context.

Understanding Regression Model Evaluation

At the core of data science and statistical analysis lies the ability to model relationships. Regression models serve as powerful tools designed to quantify the relationship between a set of predictor (or independent) variables and a dependent (or response) variable. Whether predicting stock prices, house values, or biological responses, the ultimate goal of fitting such a model is to minimize the discrepancy between the values predicted by the model and the actual observed values in the dataset. This minimization process is essential to ensure high predictive fidelity.

When we successfully train a regression model, a critical subsequent step is to quantify its effectiveness. We must understand precisely how well the model generalizes and how accurately it uses the predictor variables to forecast the response variable. This evaluation relies on calculating residuals--the differences between the observed data points and the model's predictions. These residuals are then aggregated into single summary statistics, such as MAE or RMSE, to provide a concise measure of overall model fit. The lower the value of these metrics, the better the fit,

indicating that the model's predictions are, on average, closer to the true outcomes.

The choice of which error metric to use is not arbitrary; it reflects a fundamental decision about the importance of different types of errors. The selection impacts model training, especially when regularization or optimization techniques are employed that utilize these metrics directly or indirectly. Therefore, a thorough understanding of the mathematical construction of MAE and RMSE is essential before committing to one for model assessment or optimization. These metrics are the yardsticks by which we measure the success of a predictive system in achieving high forecasting accuracy.

Defining Mean Absolute Error (MAE)

The Mean Absolute Error (MAE) is perhaps the most straightforward and intuitive measure of prediction accuracy. It represents the average magnitude of the errors in a set of predictions, without considering their direction. Essentially, MAE calculates the arithmetic mean of the absolute differences between the predicted values (\hat{y}_i) and the actual observed values (y_i). This simplicity makes MAE highly interpretable: if an MAE score is 5, it signifies that, on average, the model's predictions are off by 5 units of the response variable. This direct relationship to the units of measure provides immediate practical context to stakeholders.

One of the defining characteristics of MAE is its linear nature. Since the absolute value function is used to handle the differences, the contribution of each error to the total MAE is directly proportional to the size of that error. For instance, an error of 10 contributes exactly twice as much as an error of 5. This linearity means that MAE treats all errors equally relative to their magnitude, which is highly advantageous when the distribution of errors is expected to be uniform or when the underlying penalty for mistakes is constant. This consistency makes MAE suitable for situations where the cost function is truly linear.

MAE is particularly favored in fields like financial forecasting or inventory management, where interpretability and robustness against extreme values are critical. Because MAE minimizes the median error, it is naturally resistant to the skewing effect of large outliers. If a dataset contains a few unusually large prediction errors, MAE provides a more stable and representative measure of typical error compared to metrics that square the residuals. Therefore, if data contamination or sporadic extreme events are expected, MAE often offers a more realistic assessment of typical predictive performance.

Defining Root Mean Squared Error (RMSE)

The Root Mean Squared Error (RMSE) is another widely used metric, often employed interchangeably with MAE, yet possessing fundamentally different mathematical properties. RMSE is calculated by taking the square root of the average of the squared differences between predicted

and actual values. It essentially measures the standard deviation of the residuals, providing a metric of how concentrated the data points are around the line of best fit, analogous to the standard deviation of the residuals in standard regression output.

The crucial step in the RMSE calculation is the squaring of the errors. This operation provides a heavy penalty for large errors. Since the square of a large number is significantly larger than the number itself, large residuals contribute disproportionately more to the overall RMSE score than smaller residuals. For example, an error of 10 (squared error = 100) contributes 100 times more than an error of 1 (squared error = 1) to the total sum of squared errors, whereas in MAE, the ratio is simply 10 to 1. This characteristic makes RMSE sensitive to variance and skewness in the error distribution.

RMSE is often preferred when large errors are not just undesirable, but catastrophically costly. This is common in engineering applications, physics, or quality control, where a major failure is far worse than many small, negligible deviations. By emphasizing and magnifying larger mistakes, RMSE steers the model optimization process towards minimizing the variance of the errors, ensuring that extreme mistakes are strongly discouraged. While slightly less intuitive than MAE, RMSE remains a gold standard for evaluating model performance where minimizing high-magnitude errors is the primary objective and the loss function is quadratic.

Mathematical Formulation and Interpretation

To fully appreciate the distinction between these two metrics, examining their mathematical formulations is necessary. Both statistics utilize the observed value (y_i) and the predicted value (\hat{y}_i), along with the sample size (n), but the operations performed on the residuals ($y_i - \hat{y}_i$) determine their behavior. The core difference lies in the treatment of the residual magnitude before averaging.

The formulation for MAE is calculated as follows:

$$\text{MAE} = \frac{1}{n} \sum |y_i - \hat{y}_i|$$

The equation shows that MAE is simply the average of the absolute differences. This definition ensures that MAE is robust against the influence of extreme values, as a large error increases the overall average linearly. Furthermore, MAE has the desirable property of being convex, which simplifies its optimization in certain machine learning contexts, though the non-differentiability at zero can pose challenges for gradient-based methods used in deep learning.

The components of the MAE formula are defined as:

Σ is the summation symbol, denoting the sum across all observations.

y_i is the observed value for the i th observation.

\hat{y}_i is the predicted value for the i th observation.

n is the total number of observations (the sample size).

The calculation for RMSE, conversely, involves three distinct steps: squaring the differences, calculating the mean of these squared differences (which yields the Mean Squared Error, MSE), and finally taking the square root to return the metric to the original units of the response variable. The resulting value is always non-negative, where a value of zero indicates a perfect fit to the data, and it serves as the most popular metric when using Least Squares optimization.

It is calculated as:

$$\text{RMSE} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n}}$$

The RMSE formula's use of squaring and square-rooting ensures that it is sensitive to the variance in the errors. Because the square root function is applied last, RMSE maintains the units of the response variable, just like MAE, making both metrics interpretable on the same scale, though their numerical values will typically differ due to the inherent weighting of larger residuals.

The components of the RMSE formula are defined as:

Σ is the summation symbol.

\hat{y}_i is the predicted value for the i th observation.

y_i is the observed value for the i th observation.

n is the total number of observations (the sample size).

Practical Example: Calculating MAE & RMSE

To illustrate the computation and interpretation of these metrics, let us consider a scenario where a regression model is used to predict the number of points scored by a sample of ten basketball players in a specific game. We have the model's predictions and the actual observed scores. Calculating both metrics allows us to immediately see how they differ in magnitude and what that difference implies about the prediction errors generated by the model.

The following table presents the predicted points from the model versus the actual points scored by the players. Analyzing this raw data is the first step in assessing model performance:

Predicted Points (\hat{y}_i)	Actual points (y_i)
14	12
15	15
18	20
19	16
25	20
18	19
12	16
12	20
15	16
22	16

By applying the respective formulas to the residuals derived from this dataset, we arrive at the calculated error measures. Using the MAE formula (averaging the absolute differences), the MAE is determined to be **3.2**. This means that, on average across all ten players, the model's prediction was off by 3.2 points. This gives us a clear and intuitive measure of the typical deviation from the truth, often utilized when comparing different simpler models or when robust median performance is the target.

Conversely, applying the RMSE formula (squaring differences, averaging, and then taking the square root), the RMSE for this same dataset is found to be **4**. The interpretation here is slightly more complex: 4 represents the square root of the average squared difference between the predicted points scored and the actual points scored. Notice that in this initial example, the RMSE value (4) is higher than the MAE value (3.2). This disparity already hints at the presence of larger individual errors that the RMSE calculation is penalizing more heavily than the MAE calculation, even in a seemingly balanced dataset.

The Critical Difference: Sensitivity to Outliers

The most significant practical difference between MAE and RMSE lies in their sensitivity to outliers. This is directly attributable to the squaring operation in the RMSE formula. When large errors occur, RMSE increases significantly more than MAE, providing a disproportionate penalty that highlights the magnitude of these extreme prediction failures. This characteristic is often the deciding factor when selecting a metric for model optimization, particularly in domains where minimizing the standard deviation of error is crucial.

To vividly illustrate this sensitivity, consider introducing a single, clear outlier into our basketball scoring prediction example. Suppose one player was severely under-predicted, scoring 76 points when the model predicted only 22 points, creating a massive residual error of 54 points. This single data point dramatically alters the overall error profile of the dataset, representing a catastrophic prediction failure for that individual observation.

The following table highlights the impact of this extreme observation on the dataset:

Predicted Points (\hat{y}_i)	Actual points (y_i)
14	12
15	15
18	20
19	16
25	20
18	19
12	16
12	20
15	16
22	76

When recalculating the error metrics for this new dataset containing the outlier, the values shift substantially:

MAE: 8

RMSE: 16.4356

Observe the stark increase in the RMSE compared to the MAE. While MAE increased from 3.2 to 8 (an increase of 4.8), RMSE skyrocketed from 4 to approximately 16.44 (an increase of 12.44). This substantial difference demonstrates that the RMSE penalizes the single large error far more severely than the MAE. This behavior is due to the squared difference between the observed value of 76 and the predicted value of 22: $(76-22)^2 = 54^2 = 2,916$. This single large squared error dominates the summation in the numerator of the RMSE calculation, causing the metric to increase significantly and alerting the analyst to the presence of a substantial variance issue or poor performance on boundary cases.

Choosing the Right Metric for Your Use Case

The decision between MAE and RMSE should be driven by the specific context of the problem and the inherent costs associated with prediction errors. The core question to ask is: does the penalty for a prediction error increase linearly with the error magnitude, or does it increase exponentially? The answer dictates which metric best represents the business or scientific cost function.

If the cost of being "off" by 20 units is precisely twice as bad as being "off" by 10 units--representing a linear relationship between error magnitude and impact--then MAE is the appropriate metric. MAE assumes that the error distribution is symmetric around the mean and is ideal when robustness against outliers is critical, such as in fields focused purely on median performance or when data quality is known to be inconsistent. It is preferred when the primary concern is the average size of the error, regardless of whether that error is comprised of many small mistakes or a few large ones.

Conversely, if the penalty for being "off" by 20 units is significantly more than twice the penalty of being "off" by 10 units--meaning that large errors must be avoided at all costs--then RMSE is the superior choice. RMSE effectively minimizes the expected square loss, focusing the model on reducing variance and minimizing the chances of generating extremely large residuals. This makes RMSE particularly valuable in scenarios involving safety systems, infrastructure planning, or resource allocation where catastrophic failures are determined by the largest errors, and high forecasting accuracy is mission-critical.

Guidelines for Consistent Model Comparison

Regardless of whether MAE or RMSE is chosen, consistency is paramount when evaluating and comparing different predictive models. In practice, researchers and engineers typically fit several regression models to a single dataset, often employing various algorithms, feature sets, or hyperparameters. The evaluation process then involves calculating one standardized metric for every candidate model to ensure an equitable comparison of predictive power.

For example, a standard procedure might involve fitting three distinct regression models (e.g., Linear Regression, Random Forest, Gradient Boosting) and calculating the RMSE for each model exclusively. The model exhibiting the lowest RMSE value is then designated as the optimal model, based on the defined error criterion, because its predictions are determined to be closest to the actual observed values while minimizing the disproportionate impact of large errors.

It is absolutely crucial that the same error metric is applied uniformly across all models being compared. Comparing the MAE of one model against the RMSE of another is statistically meaningless, as the two metrics quantify error magnitude and penalty structure differently. Therefore, always commit to either MAE or RMSE at the outset of the evaluation phase and apply

that choice consistently to ensure a fair and rigorous comparison of model performance.

For those looking to implement these calculations, various statistical software packages provide built-in functionalities. The following resources may guide you in utilizing different tools:

Resources for Calculating MAE

The following tutorials explain how to calculate MAE using different statistical software, such as Python's Scikit-learn or R programming language packages:

Resources for Calculating RMSE

The following tutorials explain how to calculate RMSE using different statistical software, highlighting implementation details across various platforms:

ARABPSYCHOLOGY.COM