

What is Considered a Good AUC Score?

Authored by
stats writer

May 7, 2024

RECOMMENDED CITATION

stats writer (2024). *What is Considered a Good AUC Score?*. PSYCHOLOGICAL SCALES.
Retrieved from <https://scales.arabpsychology.com/?p=143606>

AUC (Area Under the Curve) score is a commonly used metric in machine learning and data analysis to evaluate the performance of a classification model. It measures the overall quality of the model's predictions by calculating the area under the receiver operating characteristic (ROC) curve. A higher AUC score indicates a better performing model with better predictive power.

Generally, an AUC score of 0.5 indicates a random and ineffective model, while a score of 1 represents a perfect model that can accurately classify all data points. Therefore, a good AUC score would typically be above 0.5, with a score of 0.7 or higher being considered acceptable and a score of 0.8 or higher being considered excellent.

However, the interpretation of a good AUC score may vary depending on the specific context and application. For instance, in medical research, an AUC score of 0.8 may be considered good, while in financial fraud detection, a score of 0.9 or higher may be required for a model to be considered effective.

In summary, a good AUC score is one that is higher than 0.5 and indicates a high level of predictive power for the given application. It is important to consider the specific context and requirements when evaluating the goodness of an AUC score.

What is Considered a Good AUC Score?

Logistic Regression is a method that we use to fit a regression model when the response variable is binary.

To assess how well a logistic regression model fits a dataset, we can look at the following two metrics:

Sensitivity: The probability that the model predicts a positive outcome for an observation when indeed the outcome is positive. This is also called the "true positive rate."**Specificity:** The probability that the model predicts a negative outcome for an observation when

indeed the outcome is negative. This is also called the "true negative rate."

One way to visualize these two metrics is by creating a ROC curve, which stands for "receiver operating characteristic" curve.

This is a plot that displays the sensitivity along the y-axis and (1 - specificity) along the x-axis.

One way to quantify how well the logistic regression model does at classifying data is to calculate AUC, which stands for "area under curve."

The value for AUC ranges from 0 to 1. A model that has an AUC of 1 is able to perfectly classify observations into classes while a model that has an AUC of 0.5 does no better than a model that performs random guessing.

What is a Good AUC Score?

One question students often have about AUC is:

What is a good AUC score?

The answer:

There is no specific threshold for what is considered a good AUC score.

Obviously the higher the AUC score, the better the model is able to classify observations into classes.

And we know that a model with an AUC score of 0.5 is no better than a model that performs random guessing.

However, there is no magic number that determines if an AUC score is good or bad.

If we *must* label certain scores as good or bad, we can reference the following rule of thumb from Hosmer and Lemeshow in (p. 177):

**0.5 = No discrimination
0.5-0.7 = Poor discrimination
0.7-0.8 = Acceptable discrimination
0.8-0.9 = Excellent discrimination
>0.9 = Outstanding discrimination**

A "Good" AUC Score Varies by Industry

It's important to keep in mind that what is considered a "good" AUC score varies by industry.

For example, in medical settings researchers often

seeking AUC scores above 0.95 because the cost of being wrong is so high.

For example, if we have a logistic regression model that predicts whether or not a patient will develop cancer, the price of being wrong (incorrectly telling a patient they do not have cancer when they do) is so high that we want a model that is correctly nearly every time.

Conversely, in other industries like marketing a lower AUC score may be acceptable for a model.

For example, if we have a model that predicts whether or not a customer will be a repeat customer or not, the price of being wrong is not life-altering so a model with an AUC as low as 0.6 could still be useful.

Compare AUC Scores to the Current Model

In real-world settings, we often compare the AUC scores of new logistic regression models with the AUC score of the current model being used.

For example, suppose a business uses a logistic regression model to predict whether or not customers will be repeat customers.

If the current model has an AUC score of 0.6 and you develop a new model that has an AUC of 0.65, then the new model that you have developed will be preferable even though it only offers a slight improvement and would be considered "poor" by the standards of Hosmer and Lemeshow.

The following tutorials provide additional information on how to create and interpret ROC curves and AUC scores:

ARABPSYCHOLOGY.COM