

How to Interpret the C-Statistic in Logistic Regression

Authored by
stats writer

March 4, 2026

RECOMMENDED CITATION

stats writer (2026). *How to Interpret the C-Statistic in Logistic Regression*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=133830>

The C-statistic, also known as the concordance statistic or the area under the receiver operating characteristic curve (AUC), is a measure of predictive accuracy in a logistic regression model. It quantifies the ability of the model to correctly classify observations into their respective categories. A higher C-statistic indicates a stronger predictive power of the model, while a lower C-statistic suggests a poorer fit. It is commonly used in medical research and other fields to assess the performance of a logistic regression model and determine its effectiveness in predicting outcomes.

Interpret the C-Statistic of a Logistic Regression Model

This tutorial provides a simple explanation of how to interpret the c-statistic of a logistic regression model.

What is Logistic Regression?

Logistic Regression is a statistical method that we use to fit a regression model when the *response variable* is binary. Here are some examples of when we may use logistic regression:

We want to know how exercise, diet, and weight impact the probability of having a heart attack. The response variable is *heart attack* and it has two potential outcomes: a heart attack occurs or does not occur. We want to know how GPA, ACT score, and number of AP classes taken impact the probability of getting accepted into a particular university. The response variable is *acceptance* and it has two potential outcomes: accepted or not accepted. We want to know whether

word count and email title impact the probability that an email is spam. The response variable is *spam* and it has two potential outcomes: spam or not spam.

Note that the predictor variables can be numerical or categorical; what's important is that the response variable is binary. When this is the case, logistic regression is an appropriate model to use to explain the relationship between the predictor variables and the response variable.

How to Assess the Goodness of Fit of a Logistic Regression Model

Once we have fit a logistic regression model to a dataset, we are often interested in *how well* the model fits the data. Specifically, we are interested in how well the model is able to accurately predict positive outcomes and negative outcomes.

Sensitivity refers to the probability that the model predicts a positive outcome for an observation when indeed the outcome is positive.

Specificity refers to the probability that the model predicts a negative outcome for an observation when indeed the outcome is negative.

A logistic regression model is perfect at classifying observations if it has 100% sensitivity and 100% specificity, but in practice this almost never occurs.

Once we fit the logistic regression model, it can be used to calculate the probability that a given observation has a positive outcome, based on the values of the predictor variables.

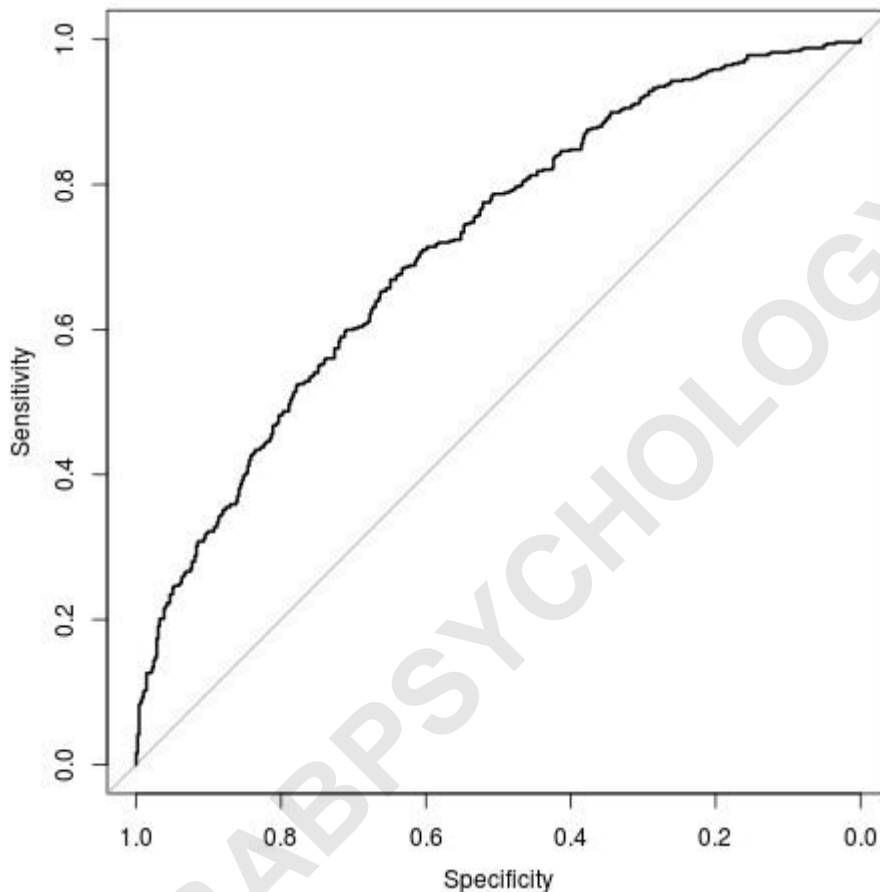
To determine if an observation should be classified as positive, we can choose a cut-point such that observations with a fitted probability above the cut-point are classified as positive and any observations with a fitted probability below the cut-point are classified as negative.

For example, suppose we choose the cut-point to be 0.5. This means that any observation with a fitted probability greater than 0.5 will be predicted to have a positive outcome, while any observation with a fitted probability less than or equal to 0.5 will be predicted to have a negative outcome.

Plotting the ROC Curve

One of the most common ways to visualize the

sensitivity vs. specificity of a model is by plotting a ROC(Receiver Operating Characteristic) curve, which is a plot of the values of sensitivity vs. 1-specificity as the value of the cut-off point moves from 0 to 1:



A model with high sensitivity and high specificity will have a ROC curve that hugs the top left corner of the plot. A model with low sensitivity and low specificity will have a curve that is close to the 45-degree diagonal line.

This means that a model with a ROC curve that hugs the top left corner of the plot would have a high area under the curve, and thus be a model that does a good job of correctly classifying outcomes. Conversely, a model with a ROC curve that hugs the 45-degree diagonal line would have a low area under the curve, and thus be a model that does a poor job of classifying outcomes.

Understanding the C-Statistic

The c-statistic, also known as the *concordance statistic*, is equal to to the AUC (area under curve) and has the following interpretations:

A value below 0.5 indicates a poor model. A value of 0.5 indicates that the model is no better out classifying outcomes than random chance. The closer the value is to 1, the better the model is at correctly classifying outcomes. A value of 1 means that the model is perfect at classifying outcomes.

Thus, a c-statistic gives us an idea about how good a model is at correctly classifying outcomes.

In a clinical setting, it's possible to calculate the c-

statistic by taking all possible pairs of individuals consisting of one individual who experienced a positive outcome and one individual who experienced a negative outcome. Then, the c-statistic can be calculated as the proportion of such pairs in which the individual who experienced a positive outcome had a higher predicted probability of experiencing the outcome than the individual who did not experience the positive outcome.

For example, suppose we fit a logistic regression model using predictor variables *age* and *blood pressure* to predict the likelihood of a heart attack.

To find the c-statistic of the model, we could identify all possible pairs of individuals consisting of one individual who experienced a heart attack and one individual who did not experience a heart attack. Then, the c-statistic can be calculated as the proportion of such pairs in which the individual who experienced the heart attack did indeed have a higher predicted probability of experiencing a heart attack compared to the individual who did not experience the heart attack.

Conclusion

In this article, we learned the following:

Logistic Regression is a statistical method that we use to fit a regression model when the *response variable* is binary. To assess the goodness of fit of a logistic regression model, we can look at the sensitivity and specificity, which tell us how well the model is able to classify outcomes correctly. To visualize the sensitivity and specificity, we can create a ROC curve. The AUC (area under the curve) indicates how well the model is able to classify outcomes correctly. When a ROC curve hugs the top left corner of the plot, this is an indication that the model is good at classifying outcomes correctly. The c-statistic is equal to the AUC (area under the curve), and can also be calculated by taking all possible pairs of individuals consisting of one individual who experienced a positive outcome and one individual who experienced a negative outcome. Then, the c-statistic is the proportion of such pairs in which the individual who experienced a positive outcome had a higher predicted probability of experiencing the outcome than the individual who did not experience the positive outcome. The closer a c-statistic is to 1, the

better a model is able to classify outcomes correctly.

ARABPSYCHOLOGY.COM