

How to Determine a Good AIC Value for Your Statistical Model

Authored by
stats writer

December 5, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Determine a Good AIC Value for Your Statistical Model*.
PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=105525>

The AIC, which stands for **Akaike's Information Criterion**, serves as a crucial tool in the field of statistical modeling. It is designed to measure the relative quality of a set of competing statistical models for a given dataset. Essentially, AIC helps researchers navigate the fundamental trade-off between model fit and model complexity. A primary goal in statistical analysis is to achieve **parsimony**--building a model that explains the maximum amount of variance in the data using the minimum number of parameters. This balance is critical because overly complex models, while fitting the current data perfectly, often suffer from overfitting and fail spectacularly when applied to new, unseen data. Therefore, a "good" AIC value is inherently a low one, signaling that the model provides a superior fit to the data while simultaneously maintaining a high degree of simplicity, thus enhancing its generalizability and predictive power.

The Mathematical Foundation of the Akaike Information Criterion

The Akaike information criterion (AIC) provides a standardized metric used to quantitatively compare the efficacy of different regression models. Developed by statistician Hirotugu Akaike, the calculation itself **penalizes models for increasing complexity**, ensuring that the selection process favors simpler models unless the added parameters substantially improve the goodness of fit. Understanding the formula is essential for grasping how AIC achieves this delicate balance between minimizing information loss and preserving parsimony.

The criterion is calculated using the following mathematical expression, which directly relates model complexity and fit:

$$\text{AIC} = 2K - 2\ln(L)$$

The components of this formula are defined as follows, each contributing uniquely to the final metric and reinforcing the complexity/fit trade-off:

K: This variable represents the **number of estimated parameters** (or degrees of freedom) within the model. This term is the primary mechanism by which AIC penalizes complexity. As the number of parameters increases, the value of $2K$ increases, driving the overall AIC score higher. This discourages the arbitrary inclusion of superfluous predictor variables that do not contribute meaningfully to the model's explanatory power.

$\ln(L)$: This is the log-likelihood of the model. The likelihood function quantifies how well the model fits the observed data, representing the probability of observing the data given the model and its estimated parameters. The natural log-likelihood ($\ln(L)$) is used for computational convenience. A

higher log-likelihood value indicates a better fit, which subsequently lowers the overall AIC score (due to the subtraction of $2\ln(L)$).

The structure of the formula demonstrates that AIC seeks to minimize the information lost when approximating reality with a chosen statistical model. The model that achieves the lowest penalty (K) while simultaneously maximizing the likelihood of the data (L) will ultimately yield the lowest AIC score and be selected as the superior choice.

Interpreting AIC: Why Absolute Values Are Meaningless

After executing and fitting several regression models to a dataset, the crucial next step is to compare their respective AIC values. The model exhibiting the **lowest AIC value** is invariably deemed the superior choice among the set of candidates being evaluated. This fundamental principle leads to a common question, particularly among students and novice practitioners: What specific numerical value constitutes a "good" AIC score?

The simple and definitive answer is that there is **no absolute value for AIC that can be considered universally "good" or "bad."** AIC is purely a metric for **relative comparison**. We use AIC solely as a ranking mechanism to determine which model, within the set being tested, provides the best balance of fit and complexity. Consequently, the absolute magnitude of the AIC score holds virtually no meaning in isolation; only the ranking matters.

To illustrate this point, consider a scenario where Model A generates an AIC value of 730.5 and Model B yields an AIC value of 456.3. In this comparison, Model B is definitively the better fitting model because 456.3 is substantially lower than 730.5. The fact that the scores are large (in the hundreds) is irrelevant; what matters is the gap between them. This comparative nature is rooted in statistical theory. As noted in standard statistical references, the absolute value of AIC is largely meaningless because it is determined by arbitrary constants inherent to the calculation. As this constant depends directly on the data, AIC can only be used to compare models fitted on identical samples. The best model from the set of plausible models being considered is therefore the one with the smallest AIC value, representing the least information loss relative to the unknown true model generating the data.

Comparing Model Efficacy Using Delta AIC

While selecting the model with the minimum AIC score (AIC_{min}) is the first step,

practitioners often require a more nuanced understanding of the strength of evidence supporting the best model compared to the others. This is achieved through the use of **Delta AIC** (Δ_i). Delta AIC is calculated by subtracting the minimum AIC observed from the AIC value of the model being evaluated: $\Delta_i = \text{AIC}_i - \text{AIC}_{\text{min}}$.

The resulting Delta AIC score provides a standardized, easily interpretable measure of the difference between models. Naturally, the best model has a $\Delta_i = 0$. However, small positive Delta AIC values indicate competing models that are still highly plausible and statistically viable alternatives. Researchers typically use established guidelines to interpret the magnitude of this difference, allowing for robust model averaging if necessary.

Δ_i between 0 and 2: Models falling into this range have substantial empirical support and should be considered strong contenders. The evidence suggesting the absolute best model is significantly superior may not be strong enough to dismiss these competing models entirely.

Δ_i between 4 and 7: Models in this range have considerably less support than the best model. They represent plausible alternatives but the evidence ratio suggests a noticeable loss of information relative to the AIC_{min} model.

Δ_i greater than 10: Models with Delta AIC scores exceeding 10 are generally considered to have virtually no support and can be safely disregarded from further consideration, as the information loss associated with these models is deemed too great.

By utilizing Delta AIC, researchers move beyond merely identifying the lowest score and gain insight into the confidence interval surrounding the selection process. This method helps ensure that all statistically viable regression models are retained for potential use or model averaging, leading to a more comprehensive analysis that accounts for uncertainty.

Limitations of AIC: The Need for Absolute Fit Metrics

While the AIC is an indispensable tool for determining which regression models perform optimally relative to one another within a defined set, it is crucial to recognize its core limitation: **it does not quantify absolute goodness-of-fit**. AIC's purpose is strictly comparative. It will always identify the best model among the candidates, even if all candidates are fundamentally poor representations of the underlying data generating process.

For instance, a particular regression model might possess the lowest AIC value among a list of potential models, but it may still be a poorly fitting model overall if the entire set of candidates fails to capture the necessary underlying systematic variation. If a researcher constructs five complex

statistical models, none of which truly captures the systematic variation in the data, AIC will select the best of those five, but that "best" model might still be grossly inadequate in explaining the phenomenon under study.

This limitation necessitates coupling AIC selection with metrics that evaluate the model's absolute performance against established statistical standards. To determine if a model truly fits a dataset well, regardless of its competitors, we must employ secondary metrics that quantify precision and bias, allowing the analyst to validate that the selected model is not only relatively superior but also substantively meaningful.

Assessing Absolute Model Fit: Utilizing Mallows' Cp

To transition from relative comparison to absolute assessment, statisticians frequently utilize metrics designed to quantify the amount of bias and precision inherent in a chosen model. One such indispensable metric, particularly for subset selection in linear regression, is **Mallows' Cp**. Mallows' Cp is essentially a metric that quantifies the amount of bias in regression models, defined as the measure of the total mean squared error (MSE) of the predictions resulting from a model, scaled relative to the MSE of the full model containing all predictors.

The core principle behind Mallows' Cp is that a chosen model should minimize both random variance and systematic bias. If a model is chosen for its simplicity (low K), but that simplicity introduces significant systematic error (bias), the model is inefficient for predictive purposes. The ideal scenario is when the model is parsimonious without sacrificing accuracy.

For a model to be considered a good absolute fit, the calculated value of Mallows' Cp should be approximately equal to K (the number of parameters in that model, including the intercept). When $C_p \approx K$, it indicates that the bias introduced by simplifying the model is minimal and acceptable. Conversely, if the C_p value significantly exceeds K , it is a strong indication that the model is highly biased and therefore a poor representation of the data, warranting rejection regardless of its AIC score.

Assessing Absolute Model Fit: Utilizing Adjusted R-squared

Another essential metric for evaluating the absolute fitness of a regression model is the **adjusted R-squared** (R^2_{adj}). Unlike the standard R-squared, which is unreliable when comparing models of different sizes because it never decreases upon adding a predictor, the adjusted R-squared introduces a necessary penalty for complexity.

The adjusted R-squared specifically measures the proportion of the variance in the response variable that can be explained by the predictor variables in the model, adjusted for the number of predictor variables in the model and the sample size. This adjustment ensures that the R^2_{adj} only increases if the newly added term improves the model's fit substantially enough to compensate for the loss of degrees of freedom.

For absolute assessment, a higher adjusted R-squared indicates a better explanatory fit. The magnitude of this metric tells the analyst the precise percentage of variability accounted for by the model. While the definition of a "good" R^2_{adj} is highly dependent on the scientific discipline, it provides crucial validation that the model selected by AIC is truly capable of explaining the observed data variation in a meaningful way.

A Step-by-Step Approach to Holistic Model Selection

Effective statistical analysis requires moving beyond simplistic reliance on any single metric. The most rigorous approach integrates the comparative power of AIC with the absolute fitness checks provided by Mallows' Cp and Adjusted R-squared. This comprehensive strategy ensures that the final model is not only superior to its competitors but also robust and statistically meaningful in an absolute sense.

The recommended procedure for analysts fitting multiple candidate statistical models follows a logical sequence:

Identify Candidate Models: Define and fit a list of plausible regression models to the data.

Relative Selection using AIC: Calculate the AIC value for each model. The model (or small subset of models) with the **lowest AIC value** is selected as the most parsimonious and relatively best-fitting option.

Quantify Absolute Fit: Then, fit this relatively best regression model to the data and calculate the secondary, absolute metrics: Mallows' Cp and the adjusted R-squared of the model.

Validation: Validate that the Mallows' Cp value is close to the number of parameters (K) and that the adjusted R-squared value is acceptable within the disciplinary context.

This comprehensive approach allows you to successfully identify the best fitting model among your candidates **and** quantify how well that model actually fits the data in absolute terms, ensuring a robust conclusion.

Conclusion: The Comparative Power of AIC

In summary, the notion of a universally "good" AIC value is statistically flawed. AIC is fundamentally a measure of **relative information loss**. Its primary objective is to balance the complexity of a statistical model (penalized by K) against its goodness of fit (rewarded by log-likelihood). The absolute numerical score is irrelevant because it is dependent on the dataset and arbitrary constants and cannot be used for comparison across different data samples.

The correct interpretation is always comparative: the model among a predefined set that possesses the **smallest AIC score is statistically the most preferred choice**. However, selecting the best model based on AIC alone is insufficient for establishing real-world validity. Researchers must subsequently validate this choice using absolute metrics like Mallows' Cp, to check for acceptable bias and parsimony, and the adjusted R-squared, to confirm explanatory power. By integrating these criteria, analysts can confidently select and validate robust, parsimonious models capable of generating accurate and reliable insights.