

How to Easily Understand Nested Models

Authored by
stats writer

December 1, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Easily Understand Nested Models*. PSYCHOLOGICAL SCALES.
Retrieved from <https://scales.arabpsychology.com/?p=103398>

In the realm of statistics and machine learning, the concept of a nested model is fundamental for robust model comparison and selection. At its core, a nested structure implies a relationship of inclusion, where one model is a restricted version of another, more complex model. This principle is often seen in hierarchical model structures where elements are organized in layers, allowing analysts to manage the complexity inherent in multivariate data analysis.

When dealing with intricate systems--whether in econometrics, computational physics, or predictive analytics--a single, monolithic model is often insufficient or overly cumbersome. Nested models address this challenge by providing a framework for creating complex representations that accurately reflect the multiple relationships between different components, such as distinct data types, variables, and theoretical objects. This hierarchical approach is vital for dissecting intricate patterns that might be obscured when attempting to rely on a single, encompassing model definition.

The utility of nested modeling extends across diverse professional domains. In statistical hypothesis testing, they enable rigorous comparison, allowing practitioners to determine if the additional complexity of a larger model is truly justified by a significant improvement in explanatory power. Understanding the mathematical and practical implications of nested relationships is essential for any practitioner focused on creating efficient, parsimonious, and highly predictive models.

A nested model is formally defined as a regression model that uses a subset of the predictor variables found in a larger, encompassing model. Essentially, if Model A can be derived from Model B by setting specific coefficients to zero, then Model A is considered to be nested within Model B. This structural dependency is what allows for direct statistical comparisons between the two forms.

To illustrate this concept clearly, consider a scenario in sports analytics where we aim to predict a basketball player's game points. We begin with a comprehensive regression model (referred to here as Model A), which incorporates four distinct predictor variables: minutes played, height, player position, and shots attempted.

The full structure of Model A is mathematically represented as:

$$\text{Points} = \beta_0 + \beta_1(\text{minutes}) + \beta_2(\text{height}) + \beta_3(\text{position}) + \beta_4(\text{shots}) + \varepsilon$$

Now, we construct a second, simpler model (Model B). This model is intentionally restricted, utilizing only a subset of the predictors from Model A--specifically, minutes played and height. The purpose of this restriction is often to test the marginal utility of the excluded variables (position and shots).

Model B, the restricted model, is defined as:

$$\text{Points} = \beta_0 + \beta_1(\text{minutes}) + \beta_2(\text{height}) + \varepsilon$$

In this setup, we conclusively state that **Model B is nested in Model A**. This designation holds true because Model B can be mathematically derived from Model A by imposing the constraint that the coefficients for the dropped variables are zero (i.e., setting $\beta_3 = 0$ and $\beta_4 = 0$). The fundamental requirement for nesting is that the simpler model is a special case of the more complex, or full, model.

It is equally important to understand what constitutes a non-nested relationship. Suppose we introduce a third structure, Model C, which seeks to predict points scored using minutes, height, and a new variable: free throws attempted.

Model C is defined by the equation:

$$\text{Points} = \beta_0 + \beta_1(\text{minutes}) + \beta_2(\text{height}) + \beta_3(\text{free throws attempted})$$

In comparing Model C against Model A (the full model), we observe that while they share some common predictor variables (minutes and height), Model C includes 'free throws attempted,' which is absent from Model A, and Model A includes 'position' and 'shots,' which are absent from Model C. Consequently, **we would not say that Model C is nested in Model A**, nor is Model A nested in Model C. The relationship is non-nested because neither model can be derived from the other simply by setting certain coefficients to zero. Direct statistical testing methods used for nested models are generally inappropriate for comparing non-nested frameworks, requiring specialized techniques like the J-test or the Cox test.

The Importance of Nested Models in Model Selection

The primary rationale for employing nested model comparisons lies in the pursuit of parsimony and efficiency in statistical modeling. When building complex predictive systems, researchers often start with a large set of potentially relevant features. However, including redundant or low-impact predictor variables unnecessarily increases model complexity, raises the risk of overfitting, and complicates interpretation. Nested modeling provides a rigorous framework for determining which variables genuinely contribute to the model's performance.

In practice, we frequently utilize nested comparisons when seeking to assess whether the inclusion of a full set of predictors significantly improves the fit to a dataset compared to a model using only a subset of those predictors. This process is crucial in fields like epidemiological research, financial forecasting, and computational linguistics, where managing model complexity against explanatory power is key to drawing reliable inferences.

For example, returning to the basketball scenario, an analyst might initially fit the **full model** (Model A), using minutes played, height, position, and shots attempted to predict points scored. While this model maximizes the information available, the analyst might harbor the suspicion that variables like 'position' and 'shots attempted' offer minimal incremental predictive value, perhaps due to high collinearity with other variables or simply having a weak fundamental relationship with the outcome variable.

To test this suspicion objectively, the analyst constructs the **nested model** (Model B), which exclusively relies on minutes played and height. By systematically comparing the goodness-of-fit metrics--such as the Residual Sum of Squares (RSS) or the likelihood function--of the two models, we can formally assess whether the marginal contribution of the excluded variables (position and shots) is statistically significant. If the difference in fit is negligible, the analyst gains confidence in dropping the extraneous variables, resulting in a simpler, more robust model that adheres to the principle of parsimony.

Mathematical Foundations: Defining Full and Restricted Models

To properly execute nested model analysis, we must formalize the concepts of the full model and the restricted model. The **full model** (often denoted as M_F) is the unrestricted model containing the complete set of K predictor variables. The restricted, or **nested model** (denoted as M_R), is defined by setting q coefficients in the full model equal to zero, where q represents the number of predictors dropped ($q < K$).

Consider the general linear regression model where Y is the dependent variable and X represents the matrix of independent variables:

Full Model (M_F):

$$Y = X\beta + \epsilon$$

The restricted model, M_R , imposes constraints on β , specifically $\beta_i = 0$ for the q variables being tested for exclusion. The fundamental requirement is that the set of explanatory variables in M_R must be a proper subset of the explanatory variables in M_F . This mathematical constraint is what validates the use of common hypothesis testing procedures like the F-test or the likelihood ratio test.

This formulation allows us to frame the comparison as a hypothesis test: are the coefficients of the dropped variables collectively equal to zero? If this hypothesis holds true, the simpler, restricted model is adequate. If we reject this hypothesis, the full model provides a significantly superior fit and the dropped variables are necessary for accurate prediction.

Formal Hypothesis Testing for Nested Models

Comparing nested models requires formalized hypothesis testing. This procedure allows us to determine if the additional complexity introduced by the full model results in a statistically meaningful improvement in fit compared to the simpler, nested model. This comparison is typically framed as testing the joint significance of the q excluded variables.

The core structure of the hypothesis test is:

H₀ (The Null Hypothesis): The full model and the nested model fit the data equally well. The constrained coefficients (the β parameters for the dropped variables) are jointly zero. Thus, for reasons of parsimony, you should **use the nested model**.

H_A (The Alternative Hypothesis): The full model fits the data significantly better than the nested model. The constrained coefficients are not jointly zero. Thus, you should **use the full model**.

This framework ensures that we only retain the more complex model if the evidence overwhelmingly suggests that the extra complexity is warranted. If the effect of the additional variables is statistically indistinguishable from zero, the simpler model is preferred.

The Likelihood Ratio Test (LRT)

One of the most robust and commonly used statistical tools for comparing nested models is the Likelihood Ratio Test (LRT). The LRT is particularly useful when dealing with models estimated using Maximum Likelihood Estimation (MLE), such as logistic or Poisson regression models. The LRT compares the fit of the restricted model (M_R) to the fit of the full model (M_F) based on their respective log-likelihood values.

The test statistic (λ) is calculated based on the difference between the maximized log-likelihoods of the two models. Under the assumptions that the null hypothesis (H_0) is true, this test statistic asymptotically follows a Chi-Square (χ^2) distribution, with degrees of freedom equal to the number of restrictions imposed (q).

The result of a likelihood ratio test is a Chi-Square test statistic and a corresponding p-value, which summarizes the probability of observing the data given that the restricted model is the true model.

If the p-value of the test is below a predetermined significance level (conventionally set at 0.05), then we have sufficient evidence to reject the null hypothesis and conclude that the full model offers a significantly better fit. Conversely, if the p-value is greater than the significance level, we fail to reject H_0 , and the simpler nested model is preferred.

The F-Test for Comparing Nested Linear Models

In the specific context of Ordinary Least Squares (OLS) linear regression models, the comparison of nested models is often conducted using the F-test, which is computationally simpler than the LRT but yields identical results under the assumption of normally distributed errors. The F-test directly compares the Residual Sum of Squares (RSS) of the two models.

The F-statistic measures the relative reduction in the error sum of squares achieved by the full model compared to the nested model, standardized by the degrees of freedom. The formula highlights that the test assesses whether the increase in the RSS when moving from M_F to M_R is large enough to be considered statistically significant.

Just like the LRT, the F-test evaluates the null hypothesis that the coefficients of the additional q variables in the full model are simultaneously zero. A large calculated F-statistic, leading to a small p-value, suggests that the full model explains significantly more variance than the nested model, justifying its complexity.

Practical Implementation and Resources

Implementing nested model comparisons is a standard operation in statistical software packages. Both the F-test and the likelihood ratio test are readily available in programming environments optimized for data science, such as R and Python, making the assessment of model parsimony a routine step in the analysis workflow. These tools handle the computation of test statistics and p-values, simplifying the process for the end-user.

Understanding how to practically apply these tests is vital for ensuring methodological rigor. The following resources provide detailed tutorials on performing these essential comparisons:

The following tutorials explain how to perform a likelihood ratio test using R and Python:

Guidance on executing the F-test for comparing nested OLS models in R, focusing on the use of the `anova()` function to compare linear regression objects.

A comprehensive walkthrough for applying the Likelihood Ratio Test in Python using statistical libraries such as `statsmodels`, particularly relevant for generalized linear models (GLMs).