

What is a Parsimonious Model?

Authored by
stats writer

April 19, 2024

RECOMMENDED CITATION

stats writer (2024). *What is a Parsimonious Model?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=137128>

A parsimonious model is a type of statistical or mathematical model that aims to explain a phenomenon with the simplest and most concise set of parameters possible. In other words, it seeks to achieve a balance between accuracy and simplicity by using the fewest number of variables necessary to accurately represent the data. This approach is based on the principle of Occam's razor, which states that the simplest explanation is often the most effective. Parsimonious models are valuable in various fields, such as economics, biology, and psychology, as they provide a more easily interpretable and generalizable understanding of complex processes. Additionally, parsimonious models can help reduce the risk of overfitting and improve the model's predictive power. Overall, a parsimonious model is a powerful tool for simplifying and understanding complex systems.

What is a Parsimonious Model?

A parsimonious model is a model that achieves a desired level of goodness of fit using as few explanatory variables as possible.

The reasoning for this type of model stems from the idea of Occam's Razor (sometimes called the "Principle of Parsimony") which says that the simplest explanation is most likely the right one.

Applied to statistics, a model that has few parameters but achieves a satisfactory level of goodness of fit should be preferred over a model that has a ton of parameters and achieves only a slightly higher level of goodness of fit.

There are two reasons for this:

1. Parsimonious models are easier to interpret and understand. Models with fewer parameters are easier to understand and explain.

2. Parsimonious models tend to have more predictive ability. Models with fewer parameters tend to perform better when applied to new data.

Consider the following two examples to illustrate these ideas.

Example 1: Parsimonious Models = Easy Interpretation

Suppose we want to build a model using a set of explanatory variables related to real estate to predict house prices. Consider the following two models along with their adjusted R-squared:

Model 1:

Equation: House price = 8,830 + 81*(sq. ft.) Adjusted R2: 0.7734

Model 2:

Equation: House price = 8,921 + 77*(sq. ft.) + 7*(sq. ft.)² - 9*(age) + 600*(rooms) + 38*(baths) Adjusted R2: 0.7823

The first model only has one explanatory variable and an adjusted R² of .7734 while the second model has five explanatory variables with only a slightly higher adjusted R².

Based on the principle of parsimony, we would prefer to use the first model because each model has roughly the same ability to explain the variation in house prices but the first model is *much* easier to understand and explain.

For example, in the first model we know that a one unit increase in square footage of a house is associated with an average house price increase of \$81. That's simple to understand and explain.

However, in the second example the coefficient estimates are much harder to interpret. For example, one additional room in the house is associated with an average house price increase of \$600, assuming that square footage, age of the house, and number of baths is held constant. That's much harder to understand and explain.

Example 2: Parsimonious Models = Better Predictions

Parsimonious models also tend to make more accurate predictions on new datasets because they're less likely to **overfit** the original dataset.

In general, models with more parameters will produce tighter fits and higher R^2 values compared to models with fewer parameters. Unfortunately, including too many parameters in a model can cause the model to fit the noise (or "randomness") of the data, rather than the true underlying relationship between the explanatory and the response variables.

This means that a highly complex model with many parameters is likely to perform poorly on a new dataset that it hasn't seen before compared to a simpler model with fewer parameters.

How to Choose a Parsimonious Model

There could be an entire course dedicated to the topic of model selection, but essentially choosing a parsimonious model comes down to choosing a model that performs best according to some metric.

Commonly used metrics that evaluate models on their performance on a training dataset *and* their number of

parameters include:

1. Akaike Information Criterion (AIC)

The AIC of a model can be calculated as:

$$\text{AIC} = -2/n * LL + 2 * k/n$$

where:

n: Number of observations in the training dataset.
LL: Log-likelihood of the model on the training dataset.
k: Number of parameters in the model.

Using this method, you can calculate the AIC of each model and then select the model with the lowest AIC value as the best model.

This approach tends to favor more complex models compared to the next method, BIC.

2. Bayesian Information Criterion (BIC)

The BIC of a model can be calculated as:

$$\text{BIC} = -2 * LL + \log(n) * k$$

where:

n: Number of observations in the training dataset.
log: The natural logarithm (with base e)
LL: Log-likelihood of the model on the training dataset.
k: Number of parameters in the model.

Using this method, you can calculate the BIC of each model and then select the model with the lowest BIC value as the best model.

This approach tends to favor models with fewer parameters compared to the AIC method.

3. Minimum Description Length (MDL)

The MDL is a way of evaluating models that comes from the field of information theory. It can be calculated as:

$$\text{MDL} = L(h) + L(D | h)$$

where:

h: The model.
D: Predictions made by the model.
L(h): Number of bits required to represent the model.
L(D | h): Number of bits required to represent the predictions

from the model on the training data.

Using this method, you can calculate the MDL of each model and then select the model with the lowest MDL value as the best model.

Depending on the type of problem you're working on, one of these methods - AIC, BIC, or MDL - may be preferred over the others as a way of selecting a parsimonious model.