

Summary of *Statistical Modeling: The Two Cultures*^{*}

Shizhe Zhang (3041882158)
shizhe_zhang@berkeley.edu

February 5, 2026

Leo Breiman argues that statistics has two main ways of working with data. The first way, which he calls the data modeling culture, assumes that data come from a specific mathematical model, like linear or logistic regression, delivering a simple and understandable picture of the data. The goal is often to estimate parameters and explain how variables relate to each other. This has been taken for granted in statistics for a long time, but Breiman points out that it can be risky because the conclusions depend heavily on the model assumptions, which are often unrealistic for complex real-world data. And the habit of using these models everywhere can lead to prejudice against other methods. It is dangerous to apply these models without checking and contemplating the scenarios.

The second way is the algorithmic modeling culture, where the data-generating process is treated as unknown. Instead of starting with a model, the goal is to find a function that predicts well. These methods are judged mainly by how accurate their predictions are on new data. Breiman argues that this culture, common in machine learning, often works better for modern, high-dimensional problems.

Another problem is that many models can fit the same data almost equally well but tell very different stories. This makes interpretation ambiguous. Ensemble methods like random forests may alleviate this problem by combining many models but they also make it harder to understand the underlying relationships. Simple models would trick us into simplifying the features and get a result easily, but we may miss the complex nature of the data. Sometimes making the model more complex can actually offer better insights.

Statistics should focus on solving real problems using whatever tools work best, instead of applying the same tools to every problem. Nowadays, we are facing more complex data and problems and the way data is generated has changed dramatically. Statisticians should be creative and energetic to embrace new methods and ideas.

^{*}<https://www.jstor.org/stable/2676681>