<https://projecteuclid.org/download/pdf_1/euclid.ss/1009213726>

**Paper Title:**

Statistical Modeling: The Two Cultures

**Authors:**

Leo Breiman

**Reviewer(s) name(s)/information/review number:**

Andy Liu

Leeson Chen

Ricky Clarke

**Summary**

The paper introduces and discusses the two different ways: data modeling and algorithmic models to reach conclusions about data. It largely bastardizes the statistical community for strictly using data modeling and its tools to solve problems revolving around data. It contrasts this with other communities who use algorithmic modeling that uses a diverse set of tools. They can handle large complex data, can be more accurate, and can be more informative than the former data modeling.

The author of the paper largely draws distinctions between the two by stating the behavior and mechanisms of each approach. He largely draws these distinctions from his experience in industry as a consultant and his academic career as a faculty for the Berkeley Statistics Department. His remedy for this problem is for statisticians to work collaboratively with other disciplines.

**Strengths and weaknesses**

**Strength: Strong points and examples**

Drawn from his experience in industry and academia, he provides strong examples for why the tools used in data modeling at the time were weak. For example, he discusses how goodness of fit tests and residual analysis were not good tools to use because of how many things they tested for. They did not reject linearity until the nonlinearity was extreme in models with higher dimensions.

**Strength: Simple**

Although parts of the paper go into the nitty-gritty parts of each approach, the author doesn’t use jargon to the point that its not understandable to someone outside of the field. This allowed me to grasp many of the distinctions he made between the two approaches.

**Weakness: Potentially outdated**

This paper was published in 2001. Many of the things that he suggested that the statistical community should use from algorithmic models seem to already be used.

**Weakness: Draws from experience**

It draws many of these conclusions from his own personal experience. His arguments would be even stronger if they were based on empirical data.

**Points and suggestions of improvement**

Draw more experiences from other colleagues in the field as opposed to just himself. This provides stronger argument for his case.

Include empirical data to make the arguments he presented stronger.

Update the paper to be more inline with current experiences in statistical academia.

**Exam-like question on this paper and its answer:**

What is the statistical research lacking in?

Statistical research is lacking in its toolset that it uses to model data. This is because many of the tools used are outdated and don’t really draw any sort of strong conclusions from data unless it’s in the extreme. This in strong contrast to techniques used in data science.