

Performing Valid Inference on Predicted Data

Adam Visokay

6th IMPRS-PHDS Annual Academy

December 4-6, 2024



MAX PLANCK INSTITUTE
FOR DEMOGRAPHIC RESEARCH



About me



University of Virginia
Bachelors in
History & Economics



Syracuse University
Masters in
Economics



University of Washington
PhD in
Sociology (current)



Research Interests

Computational Social Science Methods

Text as data

Machine Learning / AI

Rashomon Effect

Inference on Predicted Data (IPD)

Science of Science

Population Health

Mortality Estimation with Verbal Autopsy (VA)

Morbidity - Obesity

Critical Demography



Inference on Predicted Data

First, a thought experiment:

No person alive today has seen a dinosaur. So how do we draw them?



Inference on Predicted Data

First, a thought experiment:

No person alive today has seen a dinosaur. So how do we draw them?

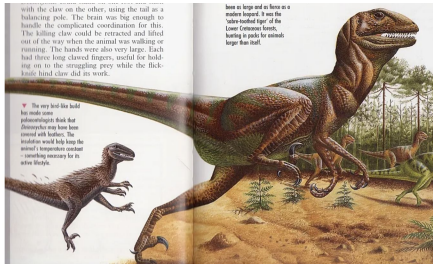


Mégalosaure restauré.

From "Les animaux d'autrefois" by Victor Meunier, 1869



Leaping Laelaps — Charles R. Knight, 1897



From The Illustrated Dinosaur Encyclopedia by Dougal Dixon, 1988



Answering Modern Scientific Questions

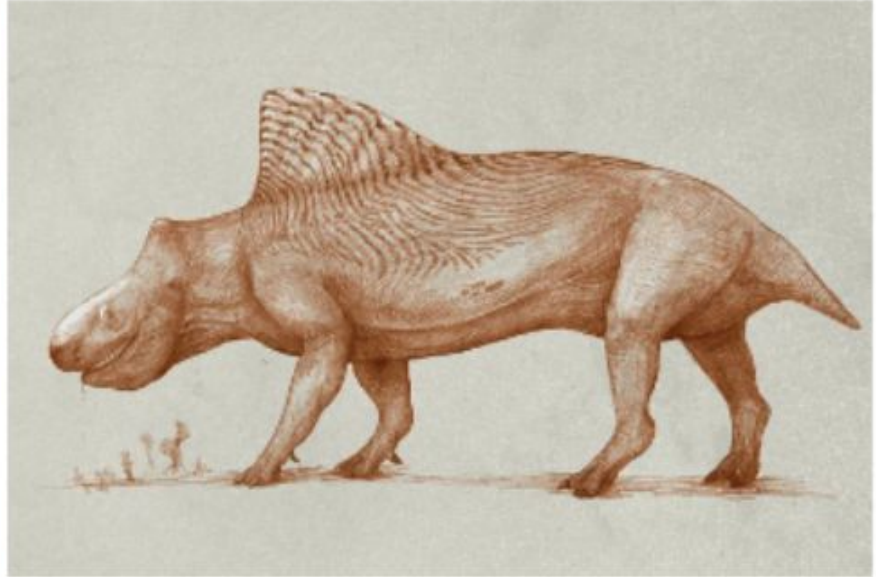


Figure 1: Artist renderings of a rhinoceroses based on limited information. Left: Albrecht Dürer's *The Rhinoceros*, woodcutting (1515); Right: C.M. Kosemen's re-imagining of a rhinoceros based on its skeleton.



Predicted Data

\$\$\$\$
(Data You Want)

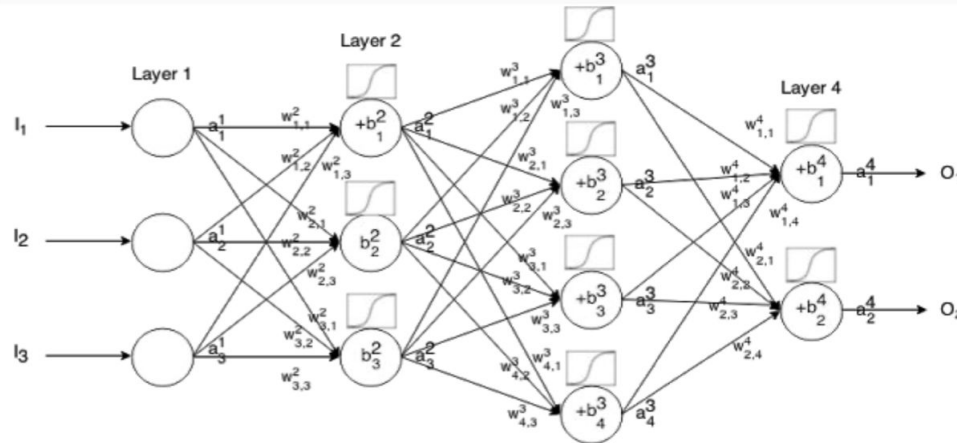


Predicted Data

\$

\$\$\$\$

$$f(\text{Data You Can Get}) = (\text{Data You Want})$$



Inference on Predicted Data

\$

\$\$\$\$

$f(\text{Data You Can Get}) = (\text{Data You Want})$



W

Inference on Predicted Data

\$

\$\$\$\$

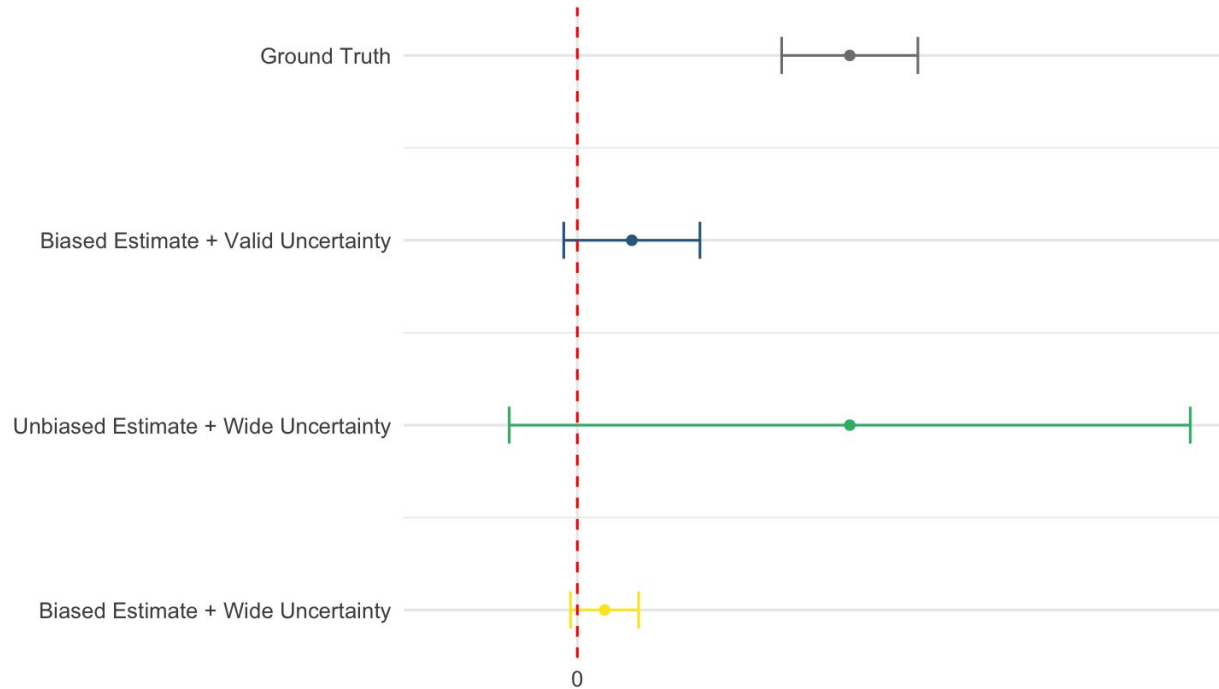
$$f(\text{Data You Can Get}) = (\text{Data You Want})$$



Valid Inference



Valid Inference



Inference on Predicted Data

Obesity $\leftrightarrow \beta$ Age



Inference on Predicted Data

Which is a better measure of obesity, BMI or DXA scanned total body fat percentage?

While BMI is cost-effective and easy to calculate, it doesn't provide the detailed body composition analysis that a DXA scan offers. If you're looking for a more comprehensive assessment of your body fat and overall health, a DXA scan might be worth the investment.

Obesity $\leftrightarrow \beta$ Age



Inference on Predicted Data

Which is a better measure of obesity, BMI or DXA scanned total body fat percentage?

While BMI is cost-effective and easy to calculate, it doesn't provide the detailed body composition analysis that a DXA scan offers. If you're looking for a more comprehensive assessment of your body fat and overall health, a DXA scan might be worth the investment.

$$\text{Obesity} \leftrightarrow \beta \text{Age}$$

$$\text{DXA} \leftrightarrow \beta_1 \text{Age}$$

$$\text{BMA} \leftrightarrow \beta_2 \text{Age}$$

$$\beta_1 \neq \beta_2$$



Examples of Predicted Data

Outcome of Interest	Cheap Prediction	Ground Truth
Cause of Death		
Obesity		
Income		
Environmental Attitude		



Examples of Predicted Data

Outcome of Interest	Cheap Prediction	Ground Truth	Covariate
Cause of Death			Age
Obesity			Age
Income			Age
Environmental Attitude			Age



Examples of Predicted Data

Outcome of Interest	Cheap Prediction	Ground Truth	Covariate
Cause of Death		Traditional Autopsy	Age
Obesity		DXA Scan Fat Percentage	Age
Income		Administrative Tax Records	Age
Environmental Attitude		Interview	Age



Examples of Predicted Data

Outcome of Interest	Cheap Prediction	Ground Truth	Covariate
Cause of Death	Verbal Autopsy Narrative	Traditional Autopsy	Age
Obesity	Body Mass Index (BMI)	DXA Scan Fat Percentage	Age
Income	Self Reported Survey	Administrative Tax Records	Age
Environmental Attitude	NLP Classification	Interview	Age



Examples of Predicted Data

Outcome of Interest	Cheap Prediction	Ground Truth	Covariate
Cause of Death	Verbal Autopsy Narrative	Traditional Autopsy	Age
Obesity	Body Mass Index (BMI)	DXA Scan Fat Percentage	Age
Income	Self Reported Survey	Administrative Tax Records	Age
Environmental Attitude	NLP Classification	Interview	Age



Thank you!

avisokay@uw.edu

<https://avisokay.github.io/>

