What is Data Science?

What's in a name?

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

What is Data Science?

First measure,
second estimate quantity,
third calculate,
fourth balance chances,
fifth succeed.

Sun Tzu, *The Art of War*, 500 BC

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

What is Data Science?

Data Scientist is the Sexiest Job of the 21st Century (Harvard Business Review)

Data Scientist is “the most sought after job in America”, according to Market Watch

Glassdoor.com has listed Data Scientist as the best job in America two years in a row

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

What is Data Science?

What is data science?

Why is data science new?

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

What is Data Science?

What is data science?

Data science is the generalization of the scientific method

Why is data science new?

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

What is Data Science?

What is data science?

Data science is the generalization of the scientific method

Why is data science new?

Data science is a "new" discipline because

1. **of an abundance of data outside of the traditional sciences**
2. **of tools to investigate these data**

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

What is Data Science?

Decompose “Data Science” into “Data” and “Science”

What is Data Science?

Decompose “Data Science” into “Data” and “Science”

- **What is the scientific method?**
- **What is the nature of data?**

What is Data Science?

What's in a name?

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

The Scientific Method

The Science in Data Science

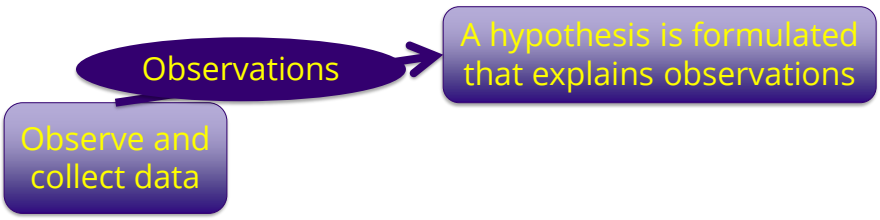
PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

The Scientific Method

Observe and collect data

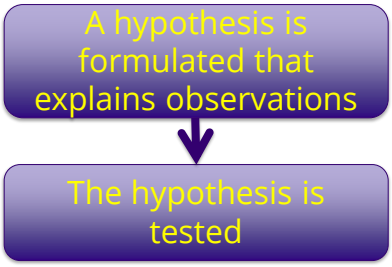
PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

The Scientific Method



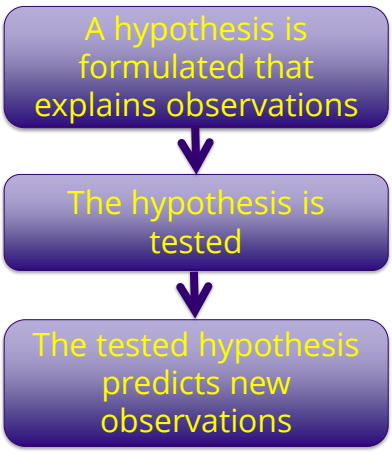
PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

The Scientific Method



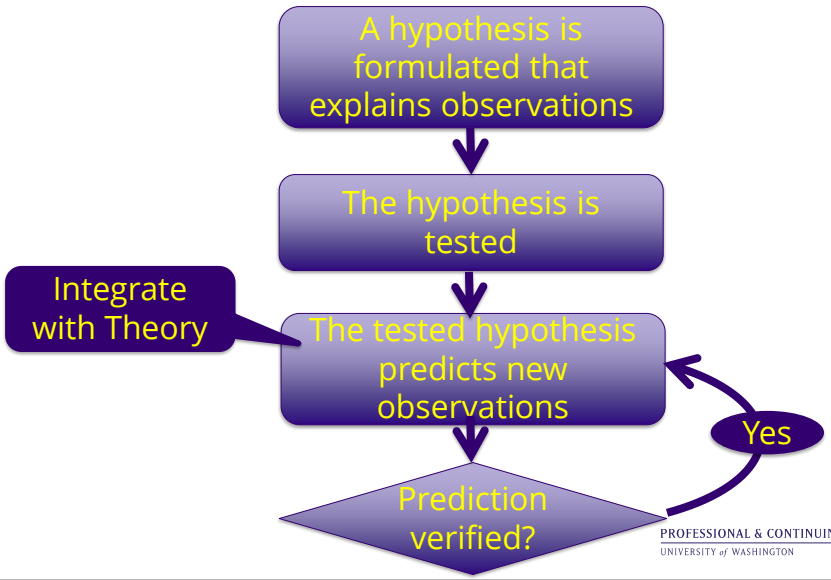
PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

The Scientific Method



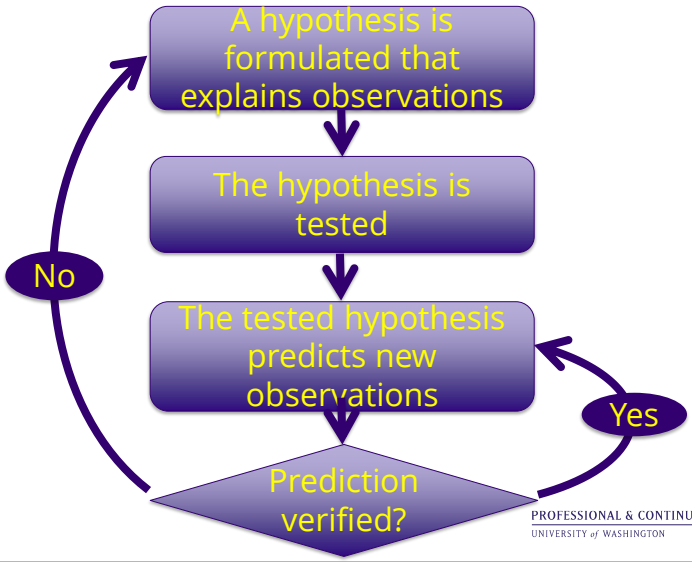
PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

The Scientific Method



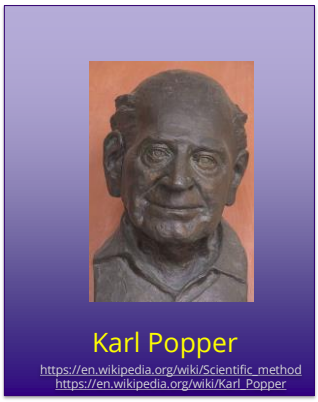
PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

The Scientific Method



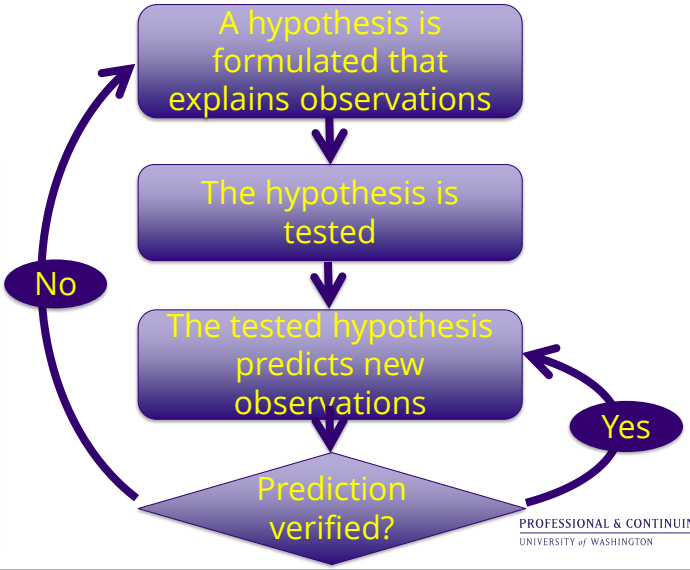
PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

The Scientific Method



Karl Popper

https://en.wikipedia.org/wiki/Scientific_method
https://en.wikipedia.org/wiki/Karl_Popper



PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

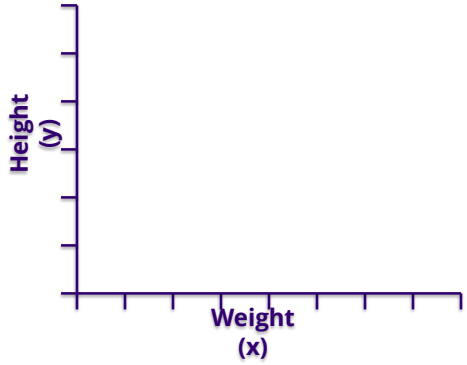
The Scientific Method

Model as a Hypothesis

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

The Scientific Method

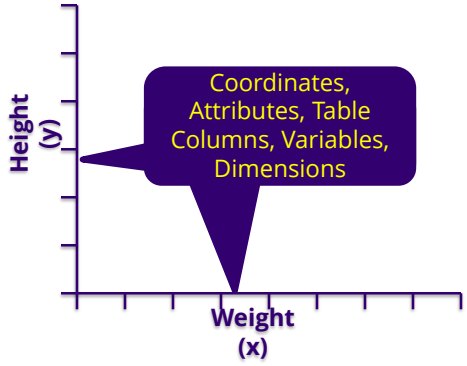
Model as a Hypothesis



PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

The Scientific Method

Model as a Hypothesis

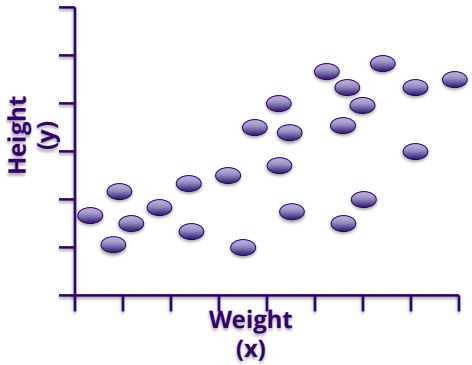


Schema, Space;

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

The Scientific Method

Model as a Hypothesis

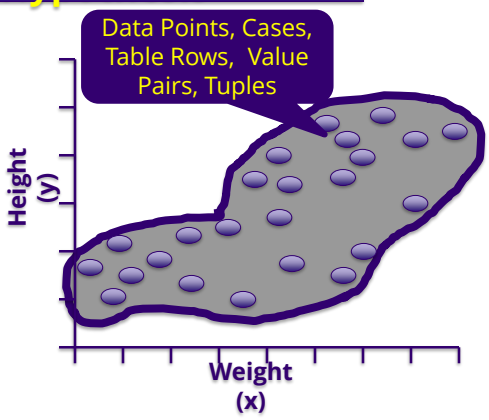


Schema, Space;
This is a two dimensional space of weight and height

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

The Scientific Method

Model as a Hypothesis

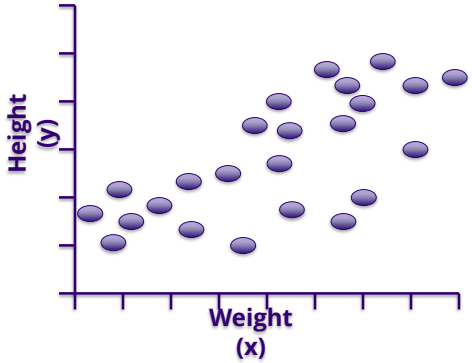


Schema, Space;
This is a two dimensional space of weight and height

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

The Scientific Method

Model as a Hypothesis

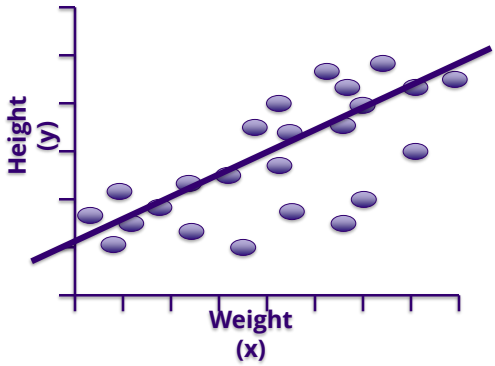


Use these data to create a model.

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

The Scientific Method

Model as a Hypothesis

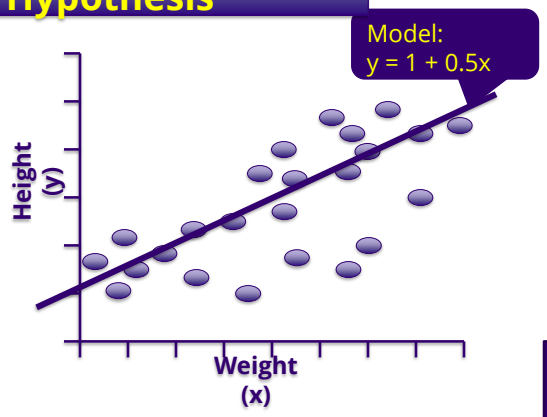


The thick line is a model that represents the ideal data

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

The Scientific Method

Model as a Hypothesis

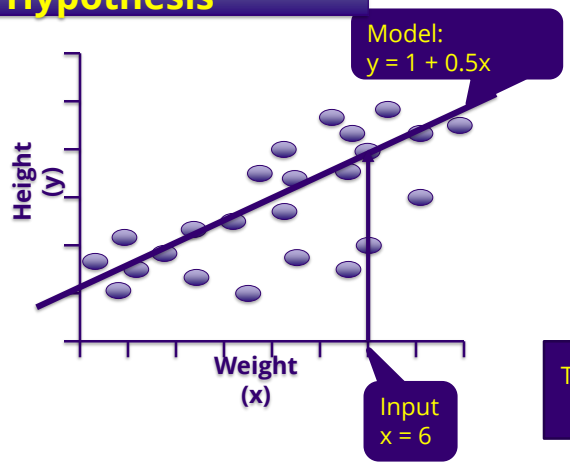


The line can be formalized. The model is the formalized statement.

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

The Scientific Method

Model as a Hypothesis

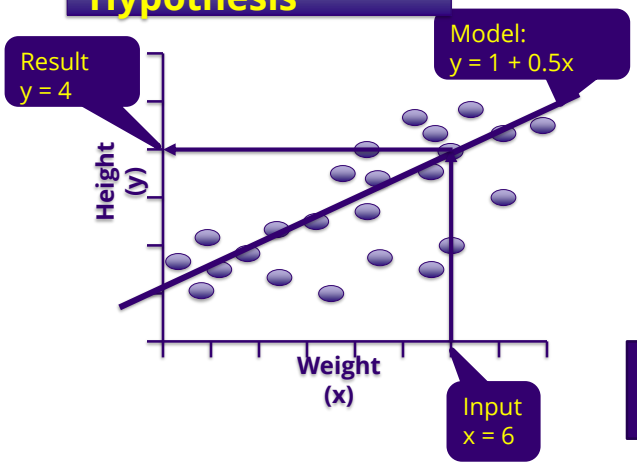


The model requires an input.

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

The Scientific Method

Model as a Hypothesis

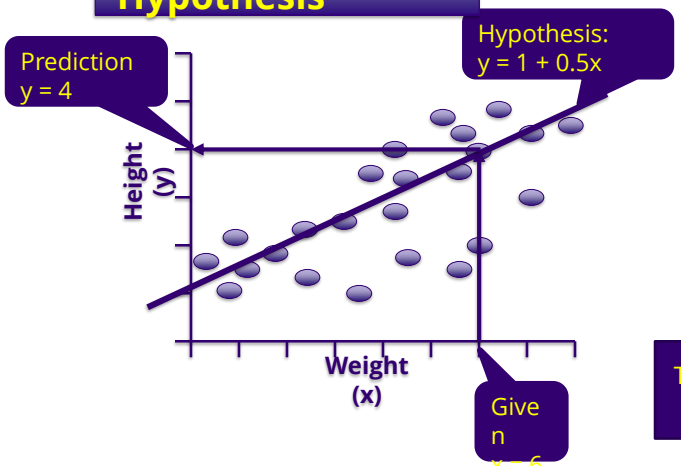


Use the model to calculate result.

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

The Scientific Method

Model as a Hypothesis

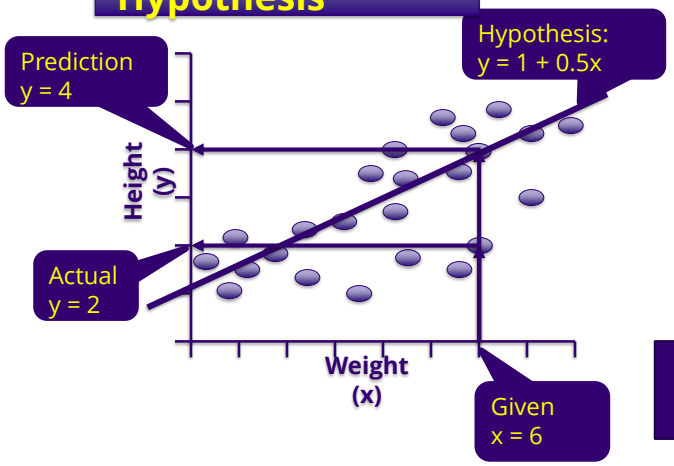


The hypothesis makes a prediction.

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

The Scientific Method

Model as a Hypothesis



Model is a Hypothesis

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

The Scientific Method

Traditional Science

Examples of data in Traditional Sciences:
- Planetary motions
- Reaction Rate of Chemicals
- Genome Sequence

Examples of models in Traditional Sciences:
- F = m*a
- Molecular Models
- Breast Cancer Probability

Data Science

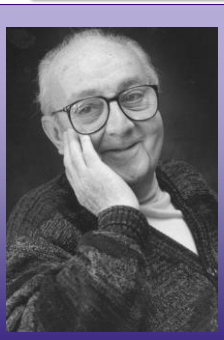
Examples of data in Data Science:
- Twitter Feeds
- Emails
- Patient Records

Examples of models in Data Science:
- Movie Sentiment
- Spam Filter
- Hospital Readmission

& CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

The Scientific Method

"All models are wrong but some are useful"



George Box

https://en.wikipedia.org/wiki/George_E._P._Box

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

The Scientific Method

Statistical Modeling: The Two Cultures

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

Leo Breiman

https://en.wikipedia.org/wiki/Leo_Brei

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

The Scientific Method

"Predictive accuracy on test sets is the criterion for how good the model is"
https://projecteuclid.org/download/pdf_1/euclid.ss/1009213726

Leo Breiman
https://en.wikipedia.org/wiki/Leo_Brei
man

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Summary: Some Terminology

>Scientific Method
  –Body of techniques
>Model
  –Based on data & method
>Hypothesis
  –Explanation of a dataset
>Falsification
  –Process to disprove hypothesis
>Theory
  –Well-tested hypothesis
>Karl Popper
  –Predictive hypotheses
>George Box
  –Some models are useful
>Leo Breiman
  –2 cultures in statistics

The Scientific Method
The Science in Data Science
PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON