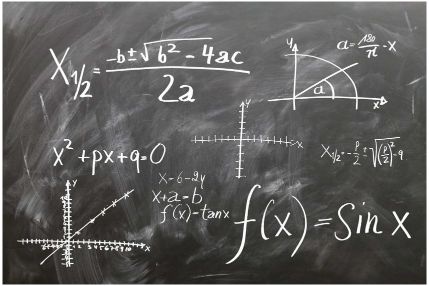
Mathematical modeling

Variables and causality

Michel Bierlaire

Introduction to transportation systems

Mathematical model



Definition

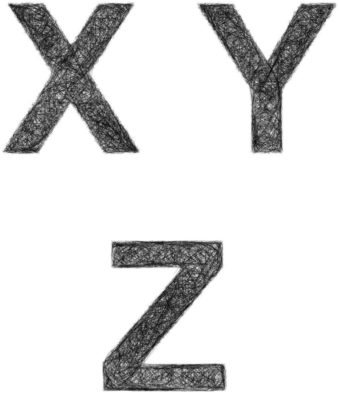
A mathematical model is a description of a system using mathematical concepts and language.

Wikipedia

Roles

- Understand.
- Predict.
- Optimize.

Modeling elements: variables



Definition

Symbol for an expression or a quantity that varies as an arbitrary object.

Wikipedia

Roles

- Capture the state of the system [traffic flow].
- Capture the decisions of the engineers [number of lanes].
- Capture the performance of the system [travel time].
- Capture external elements [weather].

Modeling elements: variables

Continuous variables

- $x \in \mathbb{R}$.
- Associated with a unit.
- Example: travel time in minutes, or seconds.

Qualitative discrete variables

- $x \in \mathcal{A}$ where $\mathcal{A}$ is a set of labels.
- Example: transportation modes $\mathcal{A} =${car as driver, car as passenger, bus, bike, train}.
- Example: level of comfort: $\mathcal{A} =$ {very comfortable, comfortable, rather comfortable, not comfortable}.

Modeling elements: variables

Binary variables

- $x \in \{0, 1\}$.
- Associated with a decision, a switch.
- Example: open a new lane or not.

Counting discrete variables

- $x \in \mathbb{N}$.
- Example: number of persons in a household.
- Note: often treated as continuous.

Modeling elements: random variables

Definition

Function:

$$X : \Omega \to \mathbb{R},$$

where Ω is a set of (random) events.

Example

Sampling individuals for a survey.

- Event: a specific individual is selected.
- Value: the income of the individual,
- or: the number of cars in the household.

Modeling elements: random variables

Notations: event

$$X = x \Longleftrightarrow \{\omega : \omega \in \Omega \text{ and } X(\omega) = x\}$$

$$X \leq x \Longleftrightarrow \{\omega : \omega \in \Omega \text{ and } X(\omega) \leq x\}$$

Discrete or continuous

- Set of possible values for X:

$$\mathcal{A} = \{x \in \mathbb{R} : (X = x) \neq \emptyset\}.$$

- If $\mathcal{A}$ is finite or countably infinite: X is discrete [number of cars].
- Otherwise, X is continuous [income].

Modeling elements: random variables

Cumulative distribution function (CDF)

$$F_X(x) = \Pr(X \leq x).$$

Property: non-decreasing

$$x < y \implies F_X(x) \leq F_X(y).$$

Probability mass function (pmf): discrete variables

$$p_X(x) = \Pr(X = x).$$

Property:

$$\sum_{x \in \mathcal{A}} p_X(x) = 1.$$

Modeling elements: random variables

Probability density function (pdf): continuous variables

$$f_X(x) = \frac{dF_X(x)}{dx}$$

Note:

$$\Pr(X = x) = 0 \text{ for strictly continuous variables.}$$

Derivation:

$$\Pr(x < X \leq x + dx) = F_X(x + dx) - F_X(x).$$

Modeling elements: random variables

Expectation: discrete

$$\mathrm{E}[X] = \sum_{x \in \mathcal{A}} x \, p_X(x).$$

Expectation: continuous

$$\mathrm{E}[X] = \int_{x \in \mathcal{A}} x \, f(x) dx.$$

Variance

$$\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2.$$

Modeling elements: random variables

Notes

- For many aspects, they can be treated as regular variables.
- As an abuse of notations, we will write

$$X \in \mathbb{R}.$$

- Similarly, we will consider vectors of random variables:

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \in \mathbb{R}^n.$$

- We will write formulas such as

$$\alpha X + \beta Y$$

Modeling elements: random variables

Linear transformations of discrete random variables

If

$$Y = \alpha X + \beta$$

then

$$p_Y(y) = p_X\left(\frac{y-\beta}{\alpha}\right)$$

Linear transformations of continuous random variables

If

$$Y = \alpha X + \beta$$

then

$$f_Y(y) = \frac{1}{|\alpha|} f_X\left(\frac{y-\beta}{\alpha}\right)$$

Mathematical model

Objective

- Explain / predict a variable(s) using other variable(s).
- Formally, we are interested in the random variable

$$Y|X = x$$

- We call Y the "dependent", "endogenous" or "explained" variable(s).
- We call X the "independent", "exogenous" or "explanatory" variable(s).

Mathematical model

Example

- X: travel time on a stretch of highway.
- Y: traffic flow on the highway.

Example

- X: number of persons in the household.
- Y: number of cars in the household.

Example

- X: weather.
- Y: number of bike trips.

Mathematical model

Objectives

- Capture causal effects.
- X and Y must be correlated.
- But correlation is not sufficient.
- For prediction, we need causality.

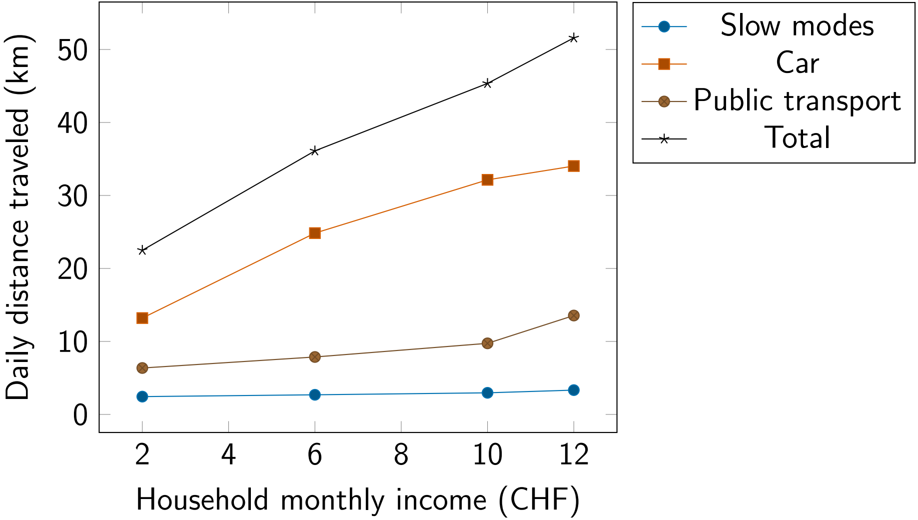
Causality and correlation

$$\text{Causality} \Rightarrow \text{correlation}$$

$$\text{Correlation} \not\Rightarrow \text{causality}$$

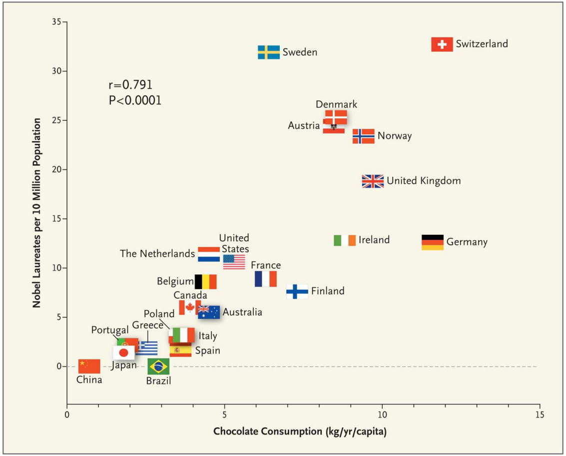
Example

Swiss Microcensus 2015. Source: ARE.



Example

Chocolate Consumption, Cognitive Function, and Nobel Laureates



Discussion: Causality? What is happening here?

Source: [Messerli, 2012]

Chocolate and Nobel prizes

Nobel | Chocolate or Chocolate | Nobel
Clear correlation
No justification for causality, in any direction
Possible explanation: wealth of the country explains both
Nobel | Wealth and Chocolate | Wealth

Spurious correlations

From `www.tylervigen.com`

- "US spending on science, space and technology", and "suicides by hanging, strangulation and suffocation".
- "Divorce rate in Maine" and "per capita consumption of margarine".
- "Per capita consumption of mozzarella cheese" and "civil engineering doctorates awarded".

Causality

Example

- X: travel time on a stretch of highway.
- Y: traffic flow on the highway.

Causality?

- $Y|X$? demand function, behavior causality.
- $X|Y$? supply function, system performance.

Example

- X: income.
- Y: distance traveled.

Causality?

- $Y|X$? makes sense.
- $X|Y$? does not make sense.

Causality

Example

- X: bus fare.
- Y: number of riders.

Causality?

- $Y|X$? demand function, behavior causality.
- $X|Y$? supply function, operator strategy.

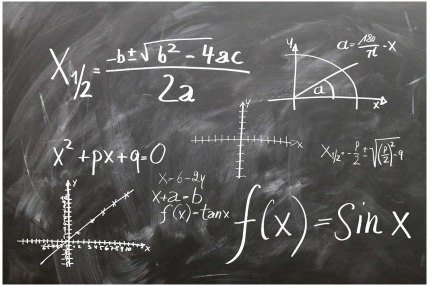
Example

- X: weather.
- Y: number of bike trips.

Causality?

- $Y|X$? makes sense.
- $X|Y$? does not make sense.

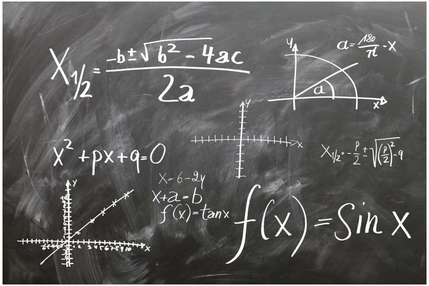
Causality



Comments

- Context dependent.
- The same variable may be exogenous or endogenous in different contexts.
- Examples: supply and demand functions.
- Importance of theoretical assumptions.

Causality



Theory

- A model relies on theory.
- Example: utility theory.
- Required for prediction, extrapolation.
- Main assumption: the causal relationship is stable over time, and over different configurations of the system.
- This is a key difference with "machine learning".

Summary

Variables

- Continuous
- Qualitative discrete
- Random

Causality

- Different than correlation.
- Context dependent.
- Relies on theory, assumptions, hypotheses.

Next

Model development

Bibliography

Messerli, F. (2012).
Chocolate consumption, cognitive function, and Nobel laureates.
The New England Journal of Medicine, 367:1562–1564.