Research questions

Christoph Hanck

Summer 2023

What is a research question anyway?

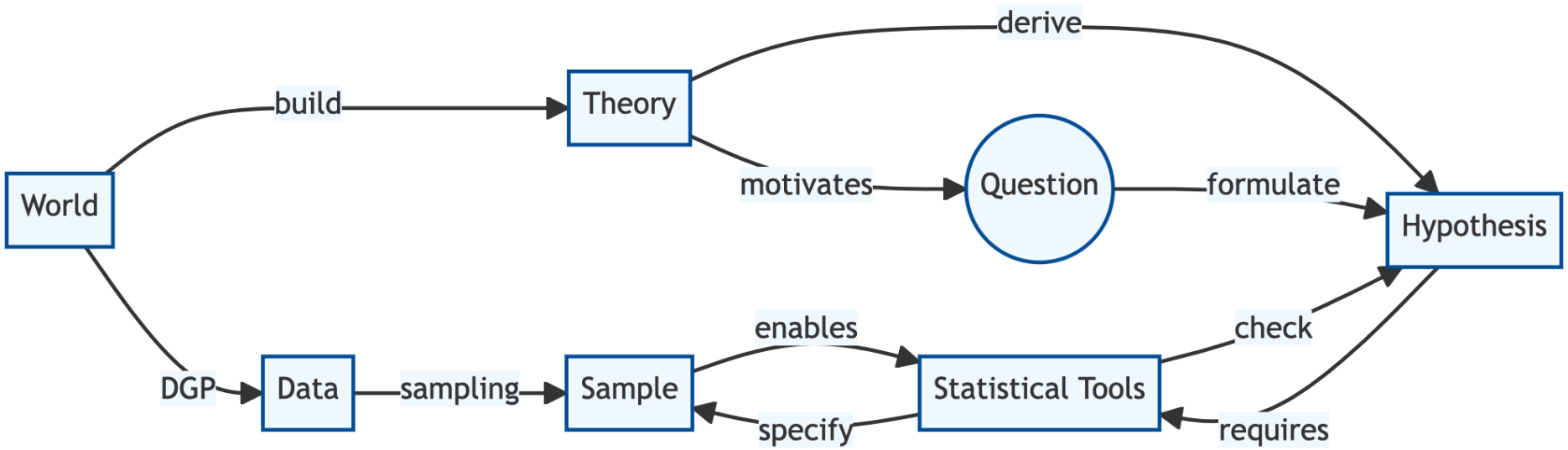
Key Koncept: Research question

A **research question** is a question

1. ... that *can be answered* and
2. ... for which having that answer will *improve our understanding of how the world works*.

A research questions takes us from a **theory** to an **hypothesis** that we can check using real world data. The data may provide **evidence** in favor or against our hypothesis.

What is a research question anyway? — ctd.



Research questions

Example: The influence of smartphone usage on children

- Does heavy smartphone usage damage children's eyesight?
- Does smartphone usage stimulate creative or intellectual activity?
- Does smartphone usage reduce children's ability to concentrate?

Why not simply look for patterns in the data?

- We are living in an information economy, surrounding us by an increasing amount of readily available data
- Why not skip hassling with theories and research questions and simply look for *patterns* in the data?
- A lot of people do — this commonly referred to as **data mining**

Definition: Data Mining

Data mining is the systematic application of computer-aided methods to discover patterns, trends or correlations in (often large and connected) data sets. Algorithms based on statistical methods are frequently used for this purpose.

Data mining

... is a useful toolkit if we care more about the *what* rather than the *why* :

- Data mining is often good at
  - ... making *predictions under stability*.
  - ... finding patterns in the data. These may foster research questions that we can examine further in other data sources.
- Data mining is less good at
  - ... helping to come up with a good theory or to improve an existing one.
  - ... producing valid inference. It has a tendency to find **false positives**.[1]

[1] A false positive is a result that incorrectly indicates the presence of a condition under examination.

Data mining — Predictions under stability

Example: Predicting die roll outcomes

- By a stable mechanism we mean that the process giving us the data does not change
- If the process is "rolling a six-sided die" data mining would based on a thousand rolls would be great at predicting that the probability of observing a 1 is 1/6
- If the process switches to "rolling a twenty-sided die", that data mining prediction will be bad: it will still predict a 1/6 chance of a 1 until it gets a lot more data.
- Probability theory will properly predict the switch to a 1/20 chance immediately

Data mining — the *what* vs. the *why*

Example on predicting ice cream sales / digression on the non-causal relation with the predictor (people wearing shorts)?

Data mining — Looking for patterns

Example: Viagra

To find a pattern can make you rich

- Pfizer scientists originally developed Viagra scientists to mitigate high blood pressure and angina.
- Famous side effects have been found using data mining: instead of coming to the data with a theory, the scientists noticed an interesting pattern in the data.

Data mining is bad as a final step

- Taking a pattern found as given would be problematic
- Pfizer did clinical studies to *see if the pattern replicated*: they verified an hypothesis derived from their research question using experimental data.

So what makes a research question a *good* one?

Simplicity

- Do not bundle a bunch of research questions into one.
- Consider potential answers — research question and theory probably do not align very well if we cannot place potential answers within the framework of our theory.

Feasibility / scalability

- A good research question can be answered using the right data.
- Data acquisition is often subject to constraints. Think time, money, data protection laws.

Suitable research design

- The research design must be suitable to investigate the derived hypothesis.
- Thoughtful consideration must be given as to which statistical method is appropriate. We will deal with this aspect in particular in this course.