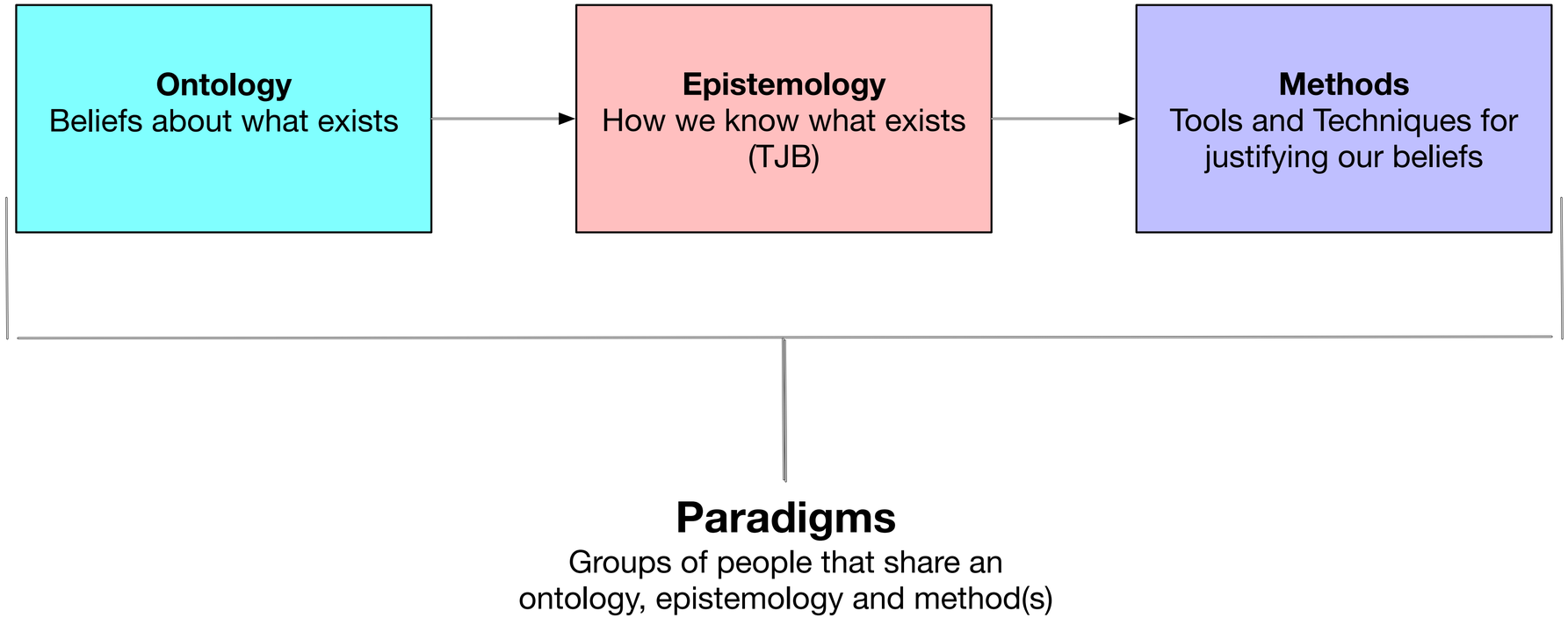
# Quantitative Data Collection Methods

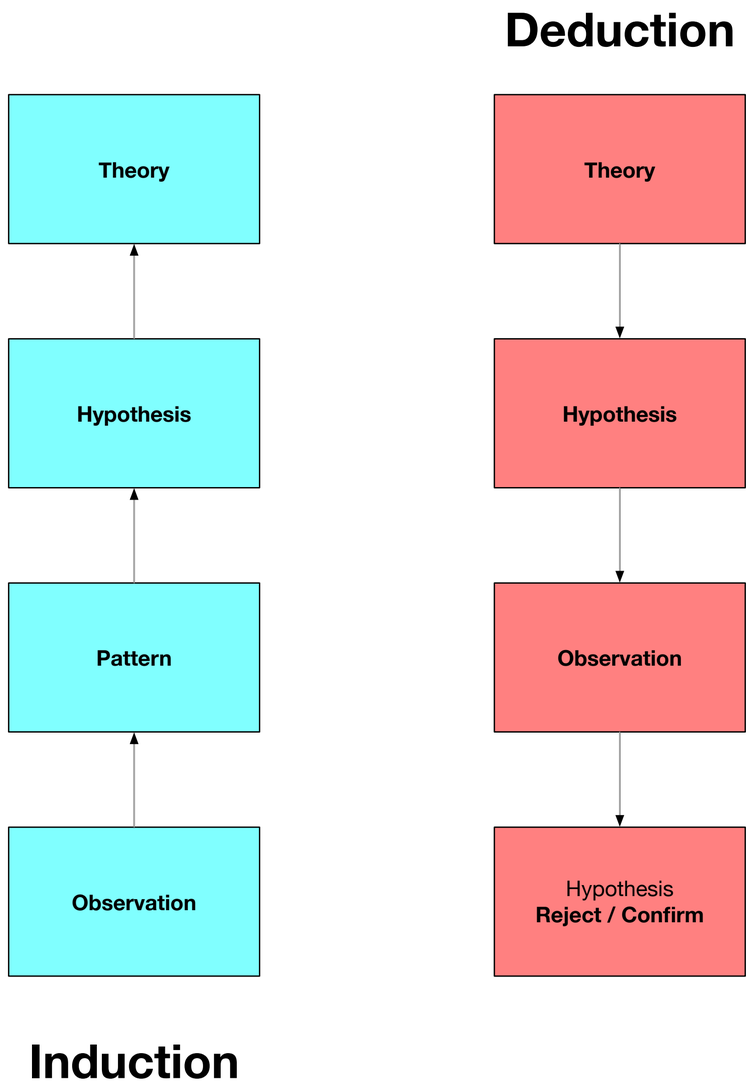
LIS 570

# Agenda

- Quant methods in relation to conceptual foundations
- Deductive logic
- Variables - Independent and Dependent
- Data Collection Methods
    - Experiment
    - Survey
    - Public / Administrative Data Reuse


**Ontology**
Beliefs about what exists
**Epistemology**
How we know what exists
(TJB)
**Methods**
Tools and Techniques for
justifying our beliefs
**Paradigms**
Groups of people that share an
ontology, epistemology and method(s)

| Research Paradigm | **Ontology**<br>Reality | **Epistemology**<br>Knowledge | **Methods**<br>Tools + Techniques |
|---|---|---|---|
| **Positivism** | There is one and only one reality / truth | Reality can be measured or inferred through logic | Surveys, data scraping, statistics<br>(usually quant) |
| **Constructivism** | Multiple truths each of which is socially constructed | Reality is interpreted and used to give meaning(s) to events, behaviors, etc. | Surveys, observation, interviews, archival analysis<br>(usually qual) |

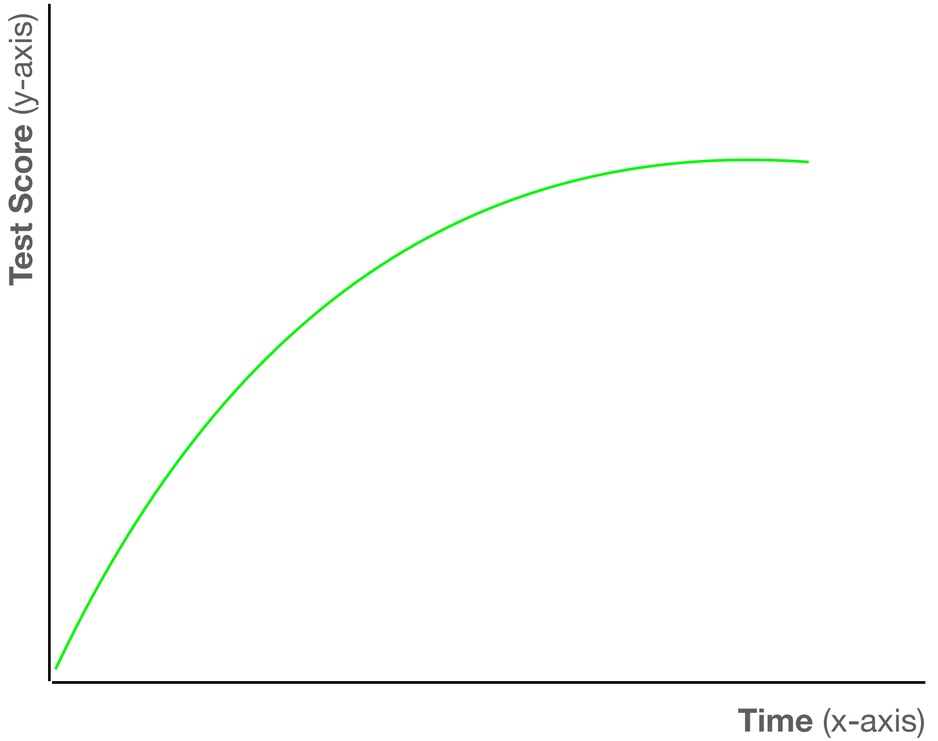

Deduction
Theory
Hypothesis
Observation
Hypothesis
Reject / Confirm
Theory
Hypothesis
Pattern
Observation
Induction

# Deductive Logic

- Measure truth value based on existing information (e.g. theory)
- Tightly control how data are collected and attempt to **reduce** bias
- State a hypothesis.
    - In statistical studies we state a Null hypothesis - meaning we expect something not to happen.
- Collect data given a purposefully designed intervention
- Accept or reject a hypothesis
    - H1: My dog has no preference for the shape of her food.

# Variables

- In quantitative research we have the ability manipulate variables so that we can ‘observe’ an effect …
    - **Independent variable -** a variable that does not depend on experimental conditions. It is *independent* of the design of a study (e.g. time)
    - **Dependent variables** - the behavior or effect that is measured. This variable is *dependent* on experimental manipulation or controls (e.g. test score)
    - **Confounding Variables** - those that are outside of our control, impact our results, and we have to acknowledge / explain in making a causal claim (e.g. setting of a test)


Test Score (y-axis)
Time (x-axis)

# Three methods

Quantitative Research

- For each method:
    - Survey
    - Experiment
    - Trace Data
- Just like with Qualitative Methods - I will give an explanation of the overarching value, sampling, instruments, data, and sensitivities of using these methods.

# Survey

## Quantitative Research

Surveys enable formal question and answering about opinions, experiences and behaviors of individuals or institutions. The self-reported data makes possible descriptive characteristics (e.g. Pew Survey on trust in government)

- **Overarching value**: Self-reporting, scale, controlled measures, replicable
- **Sampling**: Distribution-based sampling. Key to sampling errors (possible to have population level surveys)
- **Instruments:** Questionnaire (survey questions), occasionally extensible media that accompany questionnaire
- **Data are structured information that include two types of variables**
    - **Categorical variables have values that describe labels or attributes.**
        - Nominal (Nationality, gender, etc)
        - Ordinal (small, medium, large) <— Scale
    - **Numerical variables have magnitude and units, with values that carry an equal weight.**
        - Discrete - specified values (dates, size of household)
        - Continuous - any real number (temperature, height, etc)

# Experiment

## Quantitative Research

Social science experimentation is expensive, time consuming, difficult, and likely to yield small N data. But for all of these drawbacks it allows for the design of specific interventions that test behavior, preference, and motives in ways that no other method can. The design of an experiment requires at minimum a control vs an intervention - and most commonly randomized control of multiple samples.

- **Overarching value**: Maximum control, small cohort, direct measurable intervention.
- **Sampling**: Specific recruitment, screening, compensation. Often about easy to find populations (note this is a good / bad thing - WEIRD)
- **Instruments:** Administrator, task / intervention, recording
- **Data:** Very clear dependent and independent variables recorded - typically by administrator, but sometimes as log of interaction
- Example pub: Public library computer training for older adults to access high-quality Internet health information **https://doi.org/10.1016/j.lisr.2009.03.004**

# Trace / Admin Data

## Quantitative Research

Increasingly quantitive researchers can look to what are called natural experiments or opportunistic data collection based on logged events. Data that stands in for an approximate TRACE of action (clicks, comments, tweets, etc) can be used to infer motivation, behavior, etc.

- **Overarching value**: Low cost, combinable, scalable, bias-free
- **Sampling**: Usually without choice, or with minimal control (e.g. we can select a time frame to sample, a word choice, etc)
- **Instruments:** Scraping data, administrative data (API), log data - all of these require the use of data collection instruments that are
- **Data:** Structured information that requires design of analysis based on researcher identifying variables (time stamps, etc)
- Example pub: The E-book Power User in Academic and Research Libraries: Deep Log Analysis and User Customization **doi.org/10.1080/00048623.2014.885374**