Data Missingness Threats for Gathered Data
By Jinseok Kim

Threats (1): Item Nonresponse

- In gathered data, information for specific data fields can be left blank because users choose not to provide any information.
    - E.g., Twitter users may not register a profile photo and bio information about gender and birth date for privacy concerns.

Threats (2): Nondisclosure

- In gathered data, information can be missing due to data provider’s sharing policy or user’s choice of non-disclosure.
  - E.g., Twitter users can record gender on their profiles (↔ nonresponse), but Twitter blocks APIs to gather gender information.
- Many social media platforms allow users to decide which information can be shown and shared publicly.
  - E.g., gender, age, birth date, political affiliation, etc.

Threats (3): Data Astray

- In gathered data, some information may not be recorded at all because information cannot be found or is misplaced in a collection stage.
  - E.g., GPS data can have missing information because GPS signals are blocked (cars running underground) from recording.
  - E.g., Electronic health records may have blank data entry because health care workers record information into wrong entries.

Threats (4) Data Processing Errors

- Information lost during parsing.
  - E.g., forget to extract a whole data field.
- Foreign language strings that are not properly encoded can be filtered out as non-alphabetical characters.
  - E.g., 男性 (male) >> not encoded >> NULL after filtered.

Threats (5): Lost During Linkage

- Data linkage or fusion is gaining popularity to increase data value, becoming a huge area of research.
- Commonly used probabilistic matching can produce false negative (= missing) matches.
    - E.g., names of the same people recorded in different formats across data often fail to be linked.
        - ‘Jinseok Kim’ in ORCID profiles vs ‘J. S. Kim’ in publication records.

What’s Next?

- We will take a look at some examples of **data missingness threats** for gathered data with case studies.

UNIVERSITY OF MICHIGAN