The Importance of Good Research Questions for Sound Inference

Brady T. West

Revisiting Research Questions

Coming Up

- Review importance of well-formulated research questions for quality statistical inference
- See examples of inferential approaches using NHANES data to address **explicit research questions**
- Supplement each example with working Python code

Good Research Questions

- **Data are everywhere**

  → easy to find data set, import into software, run analyses
- Inferences based on those analyses tend to miss the mark in the absence of a well-formulated research question

What makes a **GOOD** research question?

Good Research Questions

Key aspects

1) What is the target population of interest?
2) Is research question descriptive or analytic?

Mean income
in a specific population?

Relationship between
income and quality of life
in a specific population?

Good Research Questions

Key aspects

3) Has **question** been asked before?
   Will new study **add knowledge** that didn't exist before?
4) Are **variables** readily **available**, **measured** appropriately, or feasible to measure using well-established tools?

Good Research Questions

Research question crafted with four properties
+ appropriate statistical procedure
= make good inferences *(related to that question)*

Absence of good research question
+ blindly running analyses
= poor insights and incorrect decisions

A Bad Question

"What is the relationship between academic performance and summer internship success?"

1. Target population? (No idea)
2. Is question descriptive or analytic? (Analytic, good)
3. Will answering question provide new knowledge? (No idea)
4. How are performance and success measured? (No idea)

A Good Question

"When considering Hispanic adults age 18+ in U.S. in 2015, what is the difference between males and females in mean systolic blood pressure?"

1. Target population clear (who, what, when, where)
2. Objectives clear (descriptive comparison of means)
3. Has question been asked before? Probably, but perhaps for other years…we are getting new knowledge!
4. Measures clear (gender, systolic blood pressure)

Good Questions Make It Easy to Choose Inferential Procedures

- **Data set** collected from a sample of Hispanic adults age 18+ in U.S. in 2015-2016 → NHANES 2015-2016
- To **compare means** between two groups (males and females) on continuous variable (systolic blood pressure)

→ Inferential procedure: independent samples t-test

Important Caveat

We will be **treating data** from NHANES
as if they come from a **simple random sample**

- Recall, complex sample design features for probability samples (*like NHANES sample*) generally need to be accounted for in inferential procedures
- More on complex sample survey analysis later!