

SURVEYS AND OBSERVATIONAL STUDIES

- Types of Survey Questions
- Mistakes in Survey Questions
- Prospective Observational Studies
- Retrospective Observational Studies
- Confounding Variables



METHODS OF DATA COLLECTION

- **Survey:** respondents are asked questions and self-report responses on various topics
- **Observational study:** researcher observes individuals and measures variables of interest but does not attempt to influence the responses
- **Experiment:** researcher deliberately imposes some condition on individuals and observes responses
 - Discussed exclusively in Lecture 4
- In observational studies and experiments, the goal is to determine if a relationship exists between an **explanatory variable** and the **response variable** by looking for associations.

EXAMPLE: IDENTIFYING METHOD OF DATA COLLECTION

- **Scenario:** A university emails a link for students, asking them to answer questions regarding the amount of time they spent on homework last week and the number of classes they skipped.
- **Question:** What method of data collection is being used?
- **Answer:** _____
 - Recipients of the email _____
 - Researcher is _____ in data collection
 - Requires _____ – researcher cannot _____

EXAMPLE: IDENTIFYING METHOD OF DATA COLLECTION

4

- **Scenario:** Recruit a group of smokers and a group of nonsmokers. Follow them over the next 10 years to see which subjects develop lung cancer.
- **Question:** What method of data collection is being used?
- **Answer:** _____
 - Looking to identify if _____ (_____ variable) is related to _____ (_____ variable)
 - Researcher collects data at _____ but does not interfere during the _____ (i.e. cannot tell subjects if they can _____)

TYPES OF SURVEY QUESTIONS

5

- **Open question:** a question where respondents can provide any answer without being forced to choose from concrete options
 - **Pro:** Provides more _____ to the statistician
 - **Con:** More _____ for the respondent to answer and for the statistician to analyze
- **Closed question:** a question where respondents have a fixed set of possible responses from which to choose; answer often indisputable
 - **Pro:** _____
 - **Con:** More _____ responses with the risk of no answer choice being _____

EXAMPLE: OPEN AND CLOSED QUESTIONS

6

- **Task:** Classify each of the following survey questions as being either open or closed.
 - **Question:** “Have you tested positive for Covid since the pandemic began?” (Choose from Yes, No, or Not sure)
 - **Answer:** _____ question – Only _____
- **Question:** “What activities did you do during summer vacation?”
- **Answer:** _____ question – _____ possible options to list so better to let the respondent provide _____

OPEN AND CLOSED QUESTIONS

7

- Many questions can be posed as either open or closed depending on how accurate you need the result to be.
 - Question regarding age may ask you input your age or to choose your age group (18-24, 25-34, ..., 65+)
 - Question regarding GPA may ask you to type your GPA or choose the correct range (2.00-2.25, 2.25-2.50, ..., 3.50+)
 - These are examples of how to categorize a quantitative variable
- Some questions may be a combination where you choose all options that apply but leave a space for “Other” where you elaborate.
 - “How have you heard about our product? Internet advertisement, TV commercial, Saw it in-store, Word of mouth, ..., Other (Please specify)

BIAS IN DESIGNING SURVEY QUESTIONS

8

- To ensure the survey responses are as representative of the population as possible, authors of surveys must be clear in their intent and avoid writing biased questions and/or answers.
- Biased survey questions often fall into one of these categories:
 - Error prone response options
 - Central tendency bias
 - Complicated question
 - Vague (or unclear) concept
 - Loaded question
 - Leading question

EXAMPLE: ERROR PRONE RESPONSE OPTIONS

9

- **Scenario:** A doctor gives his patients a survey with the following question: “How many times did you exercise the past week?”
 - Choose from: 1-2, 3-4, 5 or more
- **Question:** What is the problem with these answer choices?
- **Answer:** _____ for patients who _____
- **Question:** How can this problem be fixed?
- **Answer:** Include an option for _____

EXAMPLE: ERROR PRONE RESPONSE OPTIONS

10

- **Scenario:** A question in an exit poll after an election asked for the respondent's age with the following answer choices:
 - Choose from: 18-30, 30-40, 40-50, 50-60, 60+
- **Question:** What is the problem with these answer choices?
- **Answer:** _____
 - _____ choices for _____, _____, _____, and _____
- **Question:** How can this problem be fixed?
- **Answer:** Make the _____
 - Choose from: 18-____, 30-____, 40-____, 50-____, 60+

EXAMPLE: ERROR PRONE RESPONSE OPTIONS

11

- **Scenario:** A company sends out a survey to its employees asking: "How satisfied are you with your job?"
 - Choose from: Very satisfied, Satisfied, Somewhat satisfied
- **Question:** What is the problem with these answer choices?
- **Answer:** Assumes _____
- **Question:** How can this problem be fixed?
- **Answer:** Create _____ options
 - Choose: _____ satisfied/_____ satisfied/_____ unsatisfied/_____ unsatisfied

EXAMPLE: CENTRAL TENDENCY BIAS

12

- **Scenario:** Prior to an election, a pollster asked respondents: "On a scale from 1 to 5 with 1 being strongly oppose and 5 being strongly support, what is your opinion on using the death penalty for people convicted of first-degree murder?"
- **Question:** What response are people most likely to choose?
- **Answer:** _____
 - People without _____ tend to choose the _____ option
- **Question:** How can the bias be eliminated?
- **Answer:** Use a scale with an _____ of responses (e.g. _____)
 - Forces respondents to _____, even ever so _____

EXAMPLE: COMPLICATED QUESTION

- **Scenario:** An online survey asked respondents: “How likely are you to go out for dinner and a movie this weekend?”
 - Choose from: Very likely, Somewhat likely, Not likely
- **Question:** Why is this a complicated question?
- **Answer:** Asking about _____
 - How do you answer if you’re planning to _____?
- **Question:** How can this question be improved?
- **Answer:** Two viable options
 - Change _____ by allowing respondents to choose from _____
 - Ask: “Which of the _____ do you plan on doing this weekend?”
 - Ask _____ about how likely _____ is

EXAMPLE: VAGUE CONCEPTS

- **Scenario:** In the 1960s, a market research firm wanted to learn about stay-at-home mother’s preferred dish soap and asked the question: “What is your favorite soap?”
- **Question:** What happened when the firm got its results?
- **Answer:** Top response was _____
 - The term “soap” is _____ – could refer to _____, _____, or a _____
- **Question:** What should have been done instead?
- **Answer:** Clearly _____
 - Ask: “What is your favorite _____?”

EXAMPLE: LOADED QUESTIONS

- **Scenario:** A random sample of Pitt students is asked: “Which bars in Oakland do you enjoy patronizing?”
- **Question:** What is the problem with this question?
- **Answer:** Wording assumes that the student _____
- **Question:** How can this question be fixed?
- **Answer:** Ask if the student _____ and then ask them to _____ if the answer is _____

LEADING QUESTIONS

16

- **Scenario:** A survey asked baseball fans: “Do you believe that athletes who cheated by using steroids should be enshrined in the Hall of Fame?”
- **Question:** What is the problem with this question?
- **Answer:** Uses _____ (“_____” and “_____”)
 - Author is clearly encouraging respondents to answer _____
- **Question:** How can this question be fixed?
- **Answer:** _____ the question without the _____
 - “Should athletes who _____ steroids be _____ for the Hall of Fame?”

LOADED QUESTIONS VS. LEADING QUESTIONS

17

- **Question:** What is the primary difference between a loaded question and a leading question?
 - **Loaded Question:** “Which bars in Oakland do you enjoy patronizing?”
 - **Leading Question:** “Do you believe that athletes who cheated by using steroids should be enshrined in the Hall of Fame?”
- **Answer:**
 - **Loaded Questions:** Make an _____ about the respondent that may not be _____, making it _____ to answer _____
 - **Leading Questions:** Deliberately attempt to _____ the responses and lead the reader to answering in a _____ by using _____ words or phrases

PROS AND CONS OF SURVEYS

18

- **Question:** What are the benefits of using a survey?
- **Answer:**
 - Collect _____ in a _____
 - _____ on the part of the researcher after writing
- **Question:** What are the drawbacks?
- **Answer:**
 - Sample may exhibit _____ – some groups in the population are underrepresented in the results
 - In our example of Pitt students, students who are _____ or who are _____ are less likely to respond
 - Possibility of _____ – some people choose not to respond
 - Care must be taken to write _____ questions

TYPES OF OBSERVATIONAL STUDIES

19

- **Prospective:** establishes an outcome that the researchers watch and wait for over a period of time; once responses are collected, it is compared against other factors to identify possible relationships
 - **Pro:** Researcher involved in _____ – less prone to _____
 - **Con:** More _____ and outcome may be _____
- **Retrospective:** looks backwards to examine events that have already happened to the subject in relation to some prespecified outcome that has also already occurred
 - **Pro:** Can be performed _____
 - **Con:** May require subjects to _____ – more prone to _____

EXAMPLE: PROSPECTIVE VS. RETROSPECTIVE STUDIES

20

- **Scenario:** Recruit a group of smokers and a group of nonsmokers. Follow them over the next 10 years to see which subjects develop lung cancer.
- **Question:** What are the explanatory and response variables?
- **Answer:**
 - **Explanatory:** _____ status
 - **Response:** _____ status
- **Question:** Is this a prospective or retrospective study?
- **Answer:** _____ study
 - Response is not known until _____

EXAMPLE: PROSPECTIVE VS. RETROSPECTIVE STUDIES

21

- **Scenario:** Recruit a group of smokers and a group of nonsmokers. Follow them over the next 10 years to see which subjects develop lung cancer.
- **Question:** What are some potential problems using this design?
- **Answer:**
 - _____ – subjects _____ to have their response recorded
 - Can occur for many reasons: _____, _____, etc.
 - Subjects _____ (or _____ smoking) in the _____ of the study
 - Which _____ should they be placed into?
 - _____ – subjects are often _____ for their _____

EXAMPLE: PROSPECTIVE VS. RETROSPECTIVE STUDIES

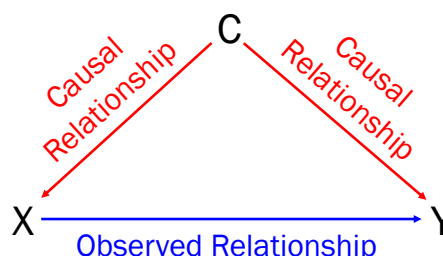
- **Scenario:** A group of people who tested positive for Covid were asked to report their vaccination status (unvaccinated, fully vaccinated, or boosted) and any symptoms they combatted
- **Question:** What are the explanatory and response variables?
- **Answer:**
 - Explanatory: _____
 - Response: _____
- **Question:** Is this a prospective or retrospective study?
- **Answer:** _____ study
 - Subjects were _____ (response is _____) and are referring to the _____ to recall both _____ and _____

EXAMPLE: PROSPECTIVE VS. RETROSPECTIVE STUDIES

- **Scenario:** A group of people who tested positive for Covid were asked to report their vaccination status (unvaccinated, fully vaccinated, or boosted) and any symptoms they combatted
- **Question:** What are some potential problems using this design?
- **Answer:**
 - _____ or _____ symptoms
 - In certain situations, finding _____ with the response could be difficult
 - e.g. Few people _____ so recruiting enough subjects to recall their symptoms may be _____

CONFOUNDING VARIABLES

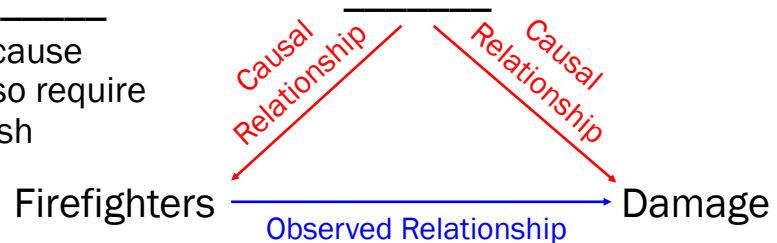
- **Confounding variable:** any variable that confuses the relationship between two variables as a result of being related to both
 - Observe what is potentially a relationship between X and Y
 - The relationship between X and Y is actually a result of the confounding variable C being causally related to both
 - X and Y have a meaningfully different relationship when C is considered



EXAMPLE: CONFOUNDING VARIABLES

- **Scenario:** A retrospective analysis of fires in a major city found that the more firefighters who responded to a fire, the more damage was caused.
- **Question:** Does having more firefighters cause more damage?
- **Answer:** _____
- **Question:** What is the confounding variable?
- **Answer:** _____

- More _____ fires will cause _____, but also require _____ to extinguish



PROS AND CONS OF OBSERVATIONAL STUDIES

- **Question:** What are the benefits of using an observational study?
- **Answer:**
 - Reveals _____ that _____ would otherwise not expose
 - Helps to frame _____
 - _____ – low risk of subjects _____
- **Question:** What are the drawbacks?
- **Answer:**
 - Cannot prove _____ – can only see if a relationship exists
 - Can be _____, causing _____ of subjects
 - May be _____