

# Design and Analysis of Sample Surveys

Andrew Gelman

Department of Statistics and Department of Political Science  
Columbia University

Class 7b: Challenges in survey measurement

- ▶ You are conducting a survey and are concerned about the possible effects of the wording of one particular question. You decide to do one of two experiments:
  - ▶ (a) Within-subject design: Put the two different wordings on the same survey form (randomizing the order of the two questions) and compare responses to the two wordings.
  - ▶ (b) Between-subject design: Randomly give one wording to half the respondents and the other wording to the other half. Compare the average responses under the two wordings.
- ▶ Give a reason why you might prefer design (a).
- ▶ Give a reason why you might prefer design (b).
- ▶ Give an example of a real survey (ideally, one that you're involved with) in which question wording is a concern.

From Tversky and Kahneman (1981):

- ▶ Problem 1: Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:
  - ▶ If Program A is adopted, 200 people will be saved.
  - ▶ If Program B is adopted, there is  $1/3$  probability that 600 people will be saved, and  $2/3$  probability that no people will be saved.

Which of the two programs would you favor?

- ▶ Problem 2:
  - ▶ If Program C is adopted 400 people will die.
  - ▶ If Program D is adopted there is  $1/3$  probability that nobody will die, and  $2/3$  probability that 600 people will die.

Which of the two programs would you favor?

# Framing: Results

- ▶ Problem 1:
  - ▶ If Program A is adopted, 200 people will be saved. (70% favor this option)
  - ▶ If Program B is adopted, there is  $1/3$  probability that 600 people will be saved, and  $2/3$  probability that no people will be saved.
- ▶ Problem 2:
  - ▶ If Program C is adopted 400 people will die.
  - ▶ If Program D is adopted there is  $1/3$  probability that nobody will die, and  $2/3$  probability that 600 people will die. (78% favor this option)

- ▶ “Choices involving gains are often risk averse and choices involving losses are often risk taking.”
- ▶ This has been known for 30 years, but economists and political scientists still talk about “utility”!
- ▶ Another example of so-called risk aversion: storable votes
  - ▶ You vote sequentially on a series of issues and you get some extra votes to use on the most important ballots
  - ▶ Strategy 1: Save your extra votes until you think you really want to use them
  - ▶ Strategy 2: Use your extra votes right away
  - ▶ Which strategy is “risk averse”?

# Measurement: Propagation of errors

- ▶ David Hemenway, “The myth of millions of annual self-defense gun uses: a case study of survey overestimates of rare events”
- ▶ Direct extrapolations from surveys suggest 2.5 million self-defense gun uses per year in the United States
- ▶ Response errors could be causing this estimate to be too high by a factor of 10
- ▶ How can this be? Discuss.

# Propagation of errors

- ▶ Direct extrapolations from surveys suggest 2.5 million self-defense gun uses per year in the United States
- ▶ Response errors could be causing this estimate to be too high by a factor of 10
- ▶ If a group truly comprises a proportion  $x$  of the population, and, in a survey, the out-group is misclassified with a rate of  $y$ , then the percentage of false positives in the population will be  $y \cdot (1 - x)$ , which, if  $x$  is small, is approximately  $y$
- ▶ If  $y$  is comparable to or larger than  $x$ , the false positives will overwhelm the true positives, and it won't work to try to crudely estimate  $x$  from a survey of the general population
- ▶ In the gun self-defense example, the most reliable estimate of  $x$  is 1 in a 1000, and so even a very low rate of misclassification will cause a direct survey estimate to be useless

# Propagation of errors

- ▶ Other examples from Hemenway:
  - ▶ “The National Rifle Association reports 3 million dues-paying members, or about 1.5% of American adults. In national random telephone surveys, however, 4–10% of respondents claim that they are dues-paying NRA members”
  - ▶ “Although Sports Illustrated reports that fewer than 3% of American households purchase the magazine, in national surveys 15% of respondents claim that they are current subscribers.”
- ▶ The discrepancies may arise because some respondents do not understand the questions or because they feel it is socially desirable to associate themselves with these organizations
- ▶ In any case, a low or moderate rate of error causes the false positives to outnumber the true positives and make the result nearly useless as an estimate of prevalence.



# Measurement: Church attendance

- ▶ Gallup: “Did you, yourself, happen to attend church or synagogue in the last seven days?”
- ▶ Hardaway, Marler, and Chaves (1993): “Our results suggest that Protestant and Catholic church attendance is roughly one-half the levels reported by Gallup.”
- ▶ Presser and Stinson (1998): “Compared to conventional interviewer-administered questions about attendance at religious services, self-administered items and time-use items . . . reduce claims of weekly attendance by about one-third. This . . . does not generally affect associations between attendance and demographic characteristics. It does, however, alter the observed trend in religious attendance over time: In contrast to the almost constant attendance rate recorded by conventional interview-administered items, approaches minimizing social desirability bias reveal that weekly attendance has declined continuously over the past three decades.”

# Measurement: Church attendance

