Chapter 4: Data Collection and Sampling Methods

El Mechry, El Koudouss

Fordham University

January 31, 2020

Types of Research Studies

- Exploratory Rsearch: identifies important variables and factors and familiarize the problem and guide subsequent research.
- Descriptive Research: evalautes the staus and trends in the topic studied.
- Causal Research: Determine whether one variable has an effect on another. This is where you go from correlation to causation. Note however, that statistical techniques alone cannot prove causality.
- Predictive Research: forcast future levels of a variable.

The Research Process

- Defining the problem,
- Deciding on the type of data required,
- Determining through what means the data will be obtained,
- Planning for the collection of data and, if necessary, selection of a sample,
- Collecting and analyzing the data,
- Drawing conclusions and reporting the findings,
- Following through with decisions that take the findings into consideration.

Survey Research: Types of surveys

- Mail Survey: In a mail survey, a mailed questionnaire is typically accompanied by a cover letter and a postage-paid return envelope for the respondent's convenience.
- Personal Interviews: An interviewer asks the respondent and documents her responses.
- Phone Interviews: The telephone interview is similar to the personal interview. It is especially useful for knowing what the respondent is doing at the time of the call (e.g., the television program, if any, being watched).

Questionnaire Design

Wording of survey questions is very important, and a number of problems can arise, including the following:

- The vocabulary level may be inappropriate for the type of person being surveyed;
- The respondent may assume a frame of reference other than the one the researcher intended;
- The question may contain "leading" words or phrases that unduly influence the response;
- The respondent may hesitate to answer a question that involves a sensitive topic.

Questionair Design Problems: Examples

- Inappropriate vocabulary level
  - Have you patronized a commercial source of cinematic entertainment within the past month?
  - Have you gone to a movie within the past month?
- Confusing frame of reference
  - Are you in better shape than you were a year ago?
  - Are you in better physical condition than you were a year ago?
- "Leading" words/phrases
  - To help maintain the quality of our schools, hospitals, and public services, do you agree that taxes should be increased next year?
  - Do you support an increase in taxes next year?

Sampling Considerations in Survey Research

Definition

A ***Compiled List*** is a sample consisting of firms or persons who are alike in some way, such as being physicians, homeowners, personal computer owners, or residents of the same area. A compiled list is passive in that you can be placed on one without having done anything.

A ***Response List*** is a sample consisting of firms or persons who have engaged in a specified behavior or activity, such as subscribing to PC Magazine, contributing to Greenpeace, buying from a mail-order catalog, or applying for an American Express credit card.

Errors in Survey Research

- Sampling Error: Sampling error occurs because a sample has been taken instead of a complete census of the population.
- Response Error: Some respondents may give false answers. They may exaggerate their income, understate their age, and so on.
- Nonresponse Error: Not everyone in the sample will participate in the survey. This is a problem if those who respond are different from those who don't.

Experimentation

In experiments, the purpose is to identify cause-and-effect relationships between variables. There are important components to an experiment:

- The treatment group: persons or objects receiving a treatment
- The control group: persons or objects receiving no treatment or a placebo
- The independent variable, or treatment
- The dependent variable, or measurement
- Extraneous variables: outside variables that are not part of the experiment, but can influence the results.

Experimentation: Example

Let T be the independent (treatment) variable, and let O denote each measurement of the dependent variable.

O_1, 1st measurement:	T, treatment:	O_2, 2nd measurement:
In 1974, 53.0% of Norwegian males smoked daily.	In 1975, Norway banned tobacco advertising.	In 1984, 41.6% of Norwegian males smoked daily.

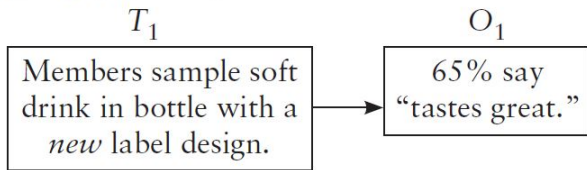Did the Norwegian ban on tobacco advertising reduce the incidence of smoking?

Validity

- Internal Validity: refers to whether T really made the difference in the measurements obtained. Was it really the advertising ban that made the difference? Could it be taxes, education, or public awareness?

- External Validity: even if T did make the difference, external validity asks whether the results can be generalized to other people or settings. Would this intervention work in USA? Could it have worked in the 50s or 60s?
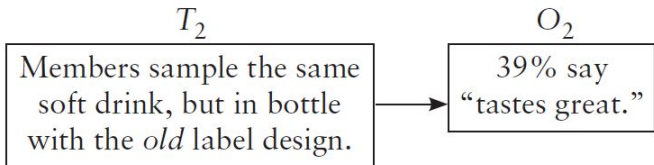
Experimentation: Example

- Experimental group ($n_1 = 100$):

$$T_1 \qquad\qquad\qquad O_1$$

| Members sample soft drink in bottle with a *new* label design. | $\rightarrow$ | 65% say "tastes great." |

- Control group ($n_2 = 100$):

$$T_2 \qquad\qquad\qquad O_2$$

| Members sample the same soft drink, but in bottle with the *old* label design. | $\rightarrow$ | 39% say "tastes great." |

Definitions

- **Population:** The set of all possible elements that could theoretically be observed or measured.
- **Sample:** A selected portion from the elements within the population, with these elements actually being measured or observed.
- **Census:** The actual measurement or observation of all possible elements from the population; this can be viewed as a "sample" that includes the entire population.
- **Parameter:** A characteristic of the population, such as the population mean μ, standard deviation σ, or population proportion π.
- **Statistic:** A characteristic of the sample, such as the sample mean $\bar{x}$, standard deviation s, or sample proportion p. In practice, this is used as an estimate of the corresponding population parameter.

Sampling Error

Sampling error is the error that occurs because a sample has been taken instead of a census.

The sample mean may differ from the true population mean simply by chance, but is just as likely to be too high as too low.

When a probability sample (discussed below) is used, the likely amount of sampling error can be statistically estimated using techniques explained in Chapter 9 and can be reduced by using a larger sample size.

Nonsampling Error

Nonsampling error, also referred to as bias, is a directional error.

For example, a weighing scale may consistently report weights that are too high.

Nonsampling errors cannot be reduced by simply increasing the size of a sample.

To reduce nonsampling error, it is necessary to take some action that will eliminate the underlying cause of the error.

Probability Sampling

In probability sampling, each person or element in the population has some (nonzero) known or calculable chance of being included in the sample. However, every person or element may not have an equal chance for inclusion.

Probability Sampling

- The Simple Random Sample: every person or element in the population has an equal chance of being included in the sample.

- The Systematic Sample: randomly select every k^{th} person, where k is chosen at random.

- Stratified Sample: the population is divided into layers, or strata; then a simple random sample of members from each stratum is selected. Strata members have the same percentage representation in the sample as they do in the population. Best when elements within strata are homogeneous.

- Cluster Sample: Randomly sample elements within some of the strata. Best when elements within strata are heterogeneous.

Nonprobability sampling

- CONVENIENCE SAMPLE: Members of such samples are chosen primarily because they are both readily available and willing to participate.
- QUOTA SAMPLE: Similar to the stratified probability sampling, but members of the various strata are not chosen through the use of a probability sampling technique.
- PURPOSIVE SAMPLE: members are chosen specifically because they're not typical of the population.
- JUDGMENT SAMPLE: members selected on the basis that the researcher believes them to be representative of the population. As a result, the representativeness of such a sample is only as good as the judgment.