

Chapter 10: Data Collection

$O = \text{truth} + \text{error} \rightarrow \text{systematic OR random (sampling) error}$

- I. Types of Error
 - a. Sampling Error
 - b. Noncoverage Error
 - c. Nonresponse error
 - d. Response error
 - e. Office error
- II. Sampling Error
 - a. Sampling error: affects projects that rely on samples drawn; difference between results obtained for a sample and the results we would have obtained had we gathered information from the population
 - b. If you don't use a probability sampling technique, it's impossible to estimate the degree of sampling error
- III. Noncoverage Error
 - a. Noncoverage error: error due to the failure to include some elements of the defined target population in the sampling frame
 - i. Sampling frame problem
 - b. Reduce by: improve the quality of the sampling frame
 - i. Use updated information
 - ii. Ex) use computerized dialing approaches to handle unlisted numbers for telephone surveys
- IV. Nonresponse error
 - a. Nonresponse error: error from failing to obtain information from some elements of the population that were selected and designated for the sample
 - b. Potential problem because it only occurs when those who respond are systematically different in some important way from those who do not respond
 - c. Two Main Sources of Nonresponse bias:
 - i. Refusals: resulting because some designated respondents refuse to participate in a study
 1. Mail & web based interviews = least affective
 - ii. Not-at-homes: arises when respondents are not at home when the interviewer calls
 1. Callbacks: attempt to reach people selected for sample by calling at different days and times
- V. Response Error
 - a. Response error: individual provides an inaccurate response, consciously or subconsciously, to a survey item
 - b. Error can stem from researcher, respondent or both
 - c. Questions for Considering Ways Response Errors can Effect Responses

- i. Does the respondent understand the question?
 1. If not:
 - a. Skip the question
 - b. Answer the question based on interpretation
 2. Pretesting the questionnaire with members of the relevant population assist with preventing response error
 - ii. Does the respondent know the answer to the question?
 1. Perform sufficient exploratory research and questionnaire pretesting to understand what population members are likely – and not likely – to know
 - iii. Is the respondent willing to provide the true answer to the question?
 1. Attitudes & emotions influence responses
 2. Perform sufficient exploratory research and questionnaire pretesting
 - iv. Is the wording of the question or the situation in which it is asked likely to bias the response?
 1. Leading questions must be avoided
 2. Don't use loaded words
 3. Watch for unstated alternative, assumed consequences, double-barreled questions
 4. Don't ask respondents for estimates
 5. Prevent by:
 - a. Keeping respondent interested and motivated
 - b. Interviewer records respondent's response by carefully writing down the person's answers to open-ended questions or checking appropriate box with closed-ended questions
 6. Interview cheating = ethically wrong
 - a. Commercial research firms follow up with respondents so ensure they have been contacted
 - b. Cheating = making up the entire interview or a few questions
- VI. Office Error
- a. Office error: error due to data editing, coding or analysis errors
 - i. Reduced by:
 1. Exercising proper controls in data processing
- VII. Total Error is Key
- a. Manage total error, not a particular type
 - b. Increasing sample size does increase sampling error, but increases other types of errors
- VIII. Calculating Response Rates
- a. Response rate: number of completed interviews with responding units divided by the number of eligible responding units in the sample

- i. Shows assessment of potential influence of nonresponse error
 1. Higher response rates = fewer problems with nonresponse bias
- ii. Response rate serves as quality indicator of data collection
 1. $RR = CI/E$
 - a. RR = Response Rate;
 - b. CI = Number of Completed Interviews with Responding Units;
 - c. E = Number of Eligible Responding Units in the Sample
- b. Web-Based and Mail Surveys
 - i. Determine usable questions completed
 - ii. Determine eligible response units
 - iii. $RR = CI/(CI + R + NAH)$
 1. RR = Response Rate;
 2. CI = Number of Completed Interviews;
 3. R = Number of Refusals;
 4. NAH = Number of Not-At-Homes
- c. Telephone Surveys (no eligibility requirement)
 - i. Categorize Attempted Contacts in 3 Groups
 1. Complete Interviews
 2. Refusals
 3. Non-at-homes
 - ii. Completed Interview: once the respondent provides answers for most of the survey items
- d. Telephone Surveys (with eligibility requirement)
 - i. Include response units that are not members of the population being studied
 1. Ex) Store wants to know shoppers' opinion on the new layout. Telephone interview those in the test-market city. Provide a screening question "Has any adult in this household visited Goldsmith's Gallery in the past 3 months?"
 - ii. Count:
 1. Completed interviews
 2. Refusal interviews
 3. Not-at-home
 4. Number of ineligible response
 - a. Help adjust for likelihood of the people who refused to take the survey or who weren't at home wouldn't have qualified anyway
 - iii. Adjust response rate to account for ineligible, by using the eligibility percentage
 1. $E\% = CI/(CI + IE)$
 - a. IE = Number of Ineligible Interviewees
 2. $RR = CI/(CI + E\% (R + NAH))$