

LECTURE 2: The Population and the Sample

Definition. An observation unit (or an element) is an object on which a measurement is taken. For example, individuals are observation units in studies of human populations.

Definition. The target population is the complete collection of all elements under study. For example, in a political poll, the target population may be all eligible voters, or all registered voters.

Definition. A sample is a subset of a sampled population.

Definition. A sampld population is the population from which the sample was taken. Note: sampled population is often different from the target population.

Definition. A sampling unit is a unit that can be selected for a sample. For example, households may be sampling units, and individuals living in these households are observation units.

Definition. A sampling frame is a list or a map or other specification of sampling units in the population from which a sample may be selected. For example, for a telephone survey, the sampling frame may be a list of all residential phone numbers in the city. For in-person interviews, the sampling frame may be a list of all street addresses.

Example. In a telephone survey of registered voters, the following structure holds

Target Population	Not reachable	Sampled Population	Sampling Frame Population
Not included in sampling frame	Refuse to respond		Not eligible for survey
	Not capable of responding		

Definition. A good sample should be as representative of the target population as possible. That is, the characteristic of interest in the target population can be estimated from a good sample with a known degree of accuracy.

Definition. A sample for which every element of a population is equally likely to be chosen is called a random sample.

Definition. If population units are sampled at a different rate than intended by an investigator, a selection bias occurs.

Definition. A convenience sample is a sample of available units. It is often biased because as a rule it doesn't include harder-to-select or non-responding units. For example, investigators took a convenience sample of adolescents to study how frequently they talk to their parents and teachers about AIDS. But adolescents who are willing to talk to the investigators are probably more likely to talk to others about AIDS.

Example. A sample of mall shopper is an example of a convenience sample. It is not representative of the entire population of consumers. The respondents at a mall are richer, for example, and more likely to be teenagers or retired. Moreover, interviewers tend to select neater, safe-looking individuals from a stream of customers.

Definition. A voluntary response sample is a sample of individuals who choose themselves by responding to general appeal. Voluntary response samples are biased because people with strong opinions, usually negative opinions, are more likely to respond. For example, a radio station may conduct call-in polls of public opinion. People who listen to the radio station and who take time (and money) to call are not representative of the entire US population.

Definition. The undercoverage of population occurs when the target population is not entirely included in the sampling frame.

Definition. The overcoverage of population occurs when the sampling frame includes some population units that are not in the target population. For example, a telephone survey may include some respondents who are younger than 18 years old, even though the target population may be individuals over 18 years old.

Definition. Nonresponse is a failure to obtain responses for all sampled units.

Definition. The response rate is the proportion of sample units for which the response is obtained.

RECOMMENDATIONS FOR QUESTIONNAIRE DESIGN

- Administer a pilot survey that tests survey questions for completeness and accuracy of wording
- Keep questions simple and clear. Avoid double negatives. For example, avoid questions like "do you favor or oppose not allowing drivers to use cell phone while driving?"
- Use specific questions rather than general ones. For example, instead of asking "did anyone attacked you in the past six months?" you may ask "has anyone attacked or threatened you in the past six months in any of the following ways: (a) with a gun, (b) with a knife, etc.
- Use multiple choice questions which responses may be quantified on a 5-point scale or a 100 % scale
- Be politically correct when wording questions (gender neutral, no reference to sexual orientation, etc.)

- Avoid questions that prompt respondents to say what you would like to hear (that is, avoid leading or loaded questions). For example, a loaded question may read “Disposal diapers account for less than 2% of the trash waste. In contrast, beverage containers and yard waste are estimated to account for 21% of trash in landfills. Given this, in your opinion, would it be fair to tax or ban disposable diapers?”
- Ask only one concept per question, that is, avoid double-barreled questions. For example, “Do you agree with Bill Clinton’s \$50 billion bailout of Mexico?”
- Pay specific attention to question order and strict logic