

2. Sampling and Measurement

- *Variable* – a characteristic that can vary in value among subjects in a sample or a population.

Types of variables

- *Categorical* (also called *qualitative*)
- *Quantitative*

(There are different statistical methods for each type)

Categorical variable – scale for measurement is a set of categories

Examples:

Racial-ethnic group (white, black, Hispanic)

Religious affiliation

Major

Political party identification (Dem., Repub., Indep.)

Vegetarian? (yes, no)

Mental health evaluation (well, mild symptom formation, moderate symptom formation, impaired)

Quantitative variable – possible values differ in magnitude

Examples:

Age, height, weight, BMI

Annual income

GPA

Time spent on Internet yesterday

Reaction time to a stimulus

(e.g., cell phone while driving experiment)

Number of “life events” in past year

Scales of measurement

For categorical variables, two types:

- *Nominal scale* – unordered categories
Preference for President, Race, Gender,
Religious affiliation, Major
Opinion items (favor vs. oppose, yes vs. no)
- *Ordinal scale* – ordered categories
Political ideology (very liberal, liberal,
moderate, conservative, very conservative)
Anxiety, stress, self esteem (high, medium, low)
Mental impairment (none, mild, moderate, severe)
Government spending on health (up, same, down)

For quantitative variables, set of possible values called an *interval scale*. (i.e., numerical interval between each possible pair of values)

Note: In practice, *ordinal* categorical variables often treated as *interval* by assigning scores

(e.g., Grades A,B,C,D,E an ordinal scale, but treated as interval if assign scores 4,3,2,1,0 to construct a GPA)

Ordering of variable types from highest to lowest level of differentiation among levels:

interval > ordinal > nominal

Another classification: **Discrete/Continuous**

Discrete variable – possible values a set of separate numbers, such as 0, 1, 2, ...

- **Example:** Number of ...
e-mail messages sent in previous day

Continuous variable – infinite continuum of possible values

Example: Amount of time spent on Internet in previous day

(In practice, distinction often blurry)

What type of variable is

1. No. of movies seen this summer (0, 1, 2, 3, 4, ...)

2. Favorite music type of (rock, jazz, folk, classical, other)

- Quantitative or categorical?
- Nominal, ordinal, or interval scale?
- Continuous or discrete?

Collecting Data in a Study

- *Sample survey*: Sample people from a population and interview them.

Example: General Social Survey. Results since 1972 at sda.berkeley.edu/GSS

(e.g., enter “heaven” and “sex” as variable names)

- *Experiment*: Compare responses of subjects under different conditions, with subjects assigned to the conditions.

Example: Harvard physicians health study – Does aspirin reduce chance of heart attack?

Randomization – the mechanism for achieving reliable data by reducing potential bias

Notation: $n = \text{sample size}$

Simple random sample: In a sample survey, each possible sample of size n has same chance of being selected.

This is an example of a *probability sampling* method – We can specify the probability any particular sample will be selected.

How to implement random sampling?

- Use “random number tables” or statistical software that can generate random numbers.
- *Sampling frame* (listing of all subjects in population) must exist to implement *simple* random sampling

Other probability sampling methods include *systematic, stratified, cluster random sampling.*
(text, pp. 21-24)

For *nonprobability sampling*, cannot specify probabilities for the possible samples.
Inferences based on them may be highly unreliable.

Example: *volunteer* samples, such as polls on the Internet, typically are severely biased.

Examples: Volunteer samples

- Lou Dobbs (CNN) asks (August 26) “Uberliberal Bill Maher says the American people are too stupid to decide whether Obama's unwritten health-care legislation is right for them, and that the president should just ram it through Congress. Do you believe that the president knows best on health care?”

Yes, I agree we need his reforms 5%

No thanks, I'll decide for myself 95%

- Text ex. (p. 20) about responses to questionnaire in the book *Women in Love*
(e.g., concluded that 70% of women married at least 5 years have extramarital affairs)

Experimental vs. observational studies

- Sample surveys are examples of **observational studies** (merely observe subjects without any experimental manipulation)
- **Experimental studies:** Researcher assigns subjects to experimental conditions.
 - Subjects should be assigned at random to the conditions (“*treatments*”)
 - Randomization “balances” treatment groups with respect to other variables that could affect response (e.g., demographic characteristics), makes it easier to assess cause and effect

Sampling error

The *sampling error* of a statistic equals the error that occurs when we use a sample statistic to predict the value of a population parameter.

Randomization protects against bias, with sampling error tending to fluctuate around 0 with predictable size

- Methods exist to predict magnitude (*margin of error*) e.g., in estimating a percentage, no more than about +3% or -3% when n about 1000 (e.g., Gallup poll)

Direction and extent of bias unknown when cannot employ randomization

Other factors besides sampling error can cause results to vary from sample to sample:

- *Sampling bias* (e.g., nonprobability sampling)
- *Response bias* (e.g., poorly worded questions, such as Lou Dobbs polls at loudobbsradio.com/surveyarchive)
- *Nonresponse bias* (undercoverage, missing data)

Read pages 19-21 of text for examples

Note: Results of surveys can depend greatly on question wording

ex. 2006 *New York Times* poll:

“Favor gasoline tax?” 12% yes

“Favor gasoline tax

to reduce U.S. dependence on foreign oil?” 55% yes

to reduce global warming?” 59% yes