Corrections: Upcoming seminars

Three "job talks"
Tue 2/16, Thu 2/18, Tue 2/23 at 3:30pm
@ UD Visitor’s Center ~210 S. College Ave.

Sociology colloquium
Thu 3/3 at 3:30pm
@ Spencer Lab Room 114, Wolf Hall Room 100

Sociology 301
Data Collection II: Sampling and Generalizability

Liying Luo
02.18

Recap

Data collection: how to collect valid data
Part I: variable and validity
Part II: sampling and generalizability

Population and Sample

a grammy.com poll "ends" (?) on Feb 15, 2016
CNN/ORC Poll: “Clinton, Sanders in a dead heat for Nevada” (Feb 17, 2016)

...based on phone interviews on Feb 10-15 with 282 Democratic likely caucus goers among 1,006 adult Nevadans: 48% supporting Clinton and 47% supporting Sanders.


Women graduates of elite liberal arts colleges “have clearly shifted away from having their family first (i.e., having at least one biological child) and a job second (i.e., having a job after 15 years of receiving their undergraduate degree but having very weak labor force attachment prior to that) toward simultaneously having both a career (i.e., very strong labor force attachment) and a family for those that graduated after 1979.”

Population and Sample

population
all of the units (e.g., individuals, cities, states, schools, prisons) about which we wish to draw conclusions about.

sample
a subset of the population the subset that we actually observe/interviewed/surveyed

size
the number of units
n = sample size; N = population size

What is the sample and sample size? Population?

the grammy.com poll
phone interviews with 2,645 “traditional” students who reported information on their graduation year and graduated from the Claremont Colleges between 1960 and 1994.

the CNN poll
phone interviews with 282 Democratic...
Sampling

- sampling: the process of selecting units from the population for inclusion in the sample
- sampling frame: a list from which the units in the population will be selected

Population, Sample, and Sampling Frame

- (target) population
- sampling frame
- sample

Why Sampling

- wish to understand some aspects of the population, but not feasible to collect information about every unit in the population
- generalizability: the ability to generalize the conclusion based on the sample to the population

Error and Bias: What Could Go Wrong?

- selection bias: the procedure used to sample units from the population is flawed
- non-response bias: the units selected to be in the sample cannot be included or refuse to be included in the sample
- measurement error (invalid measurement): our measures of sample units’ attributes are flawed
Generalizability

Evaluate the generalizability of the following...

- the grammy.com poll
- the CNN poll
- phone interviews with 282 Democratic likely caucus goers.
- the Antecol study
- 2,645 “traditional” students who reported information on their graduation year and graduated from the Claremont Colleges between 1960 and 1994.

In-Class Activity: Family Size

How many brothers and sisters do you have in your family, including yourself?

Population, Sample, and Sampling Frame

Sampling Methods

- probability sampling
  - prior knowledge about probability of selection
- non-probability sampling
  - without such prior knowledge
Sampling Methods

**probability sampling**
- the likelihood of a unit in the population to be selected for the sample
- prior knowledge about probability of selection

**non-probability sampling**
- without such prior knowledge

---

**Probability Sampling: What It Is**

random selection procedure
- every unit has a known, equal chance to be selected
- the probability of selecting a unit = the proportion of sample in the population

probability of selecting an individual in the US population?
- when sample size $n$ = the US population size?

---

**Probability Sampling: How We Do It**

**simple random sampling**

steps:
- (a) numbering each unit
- (b) generate $n$ random numbers ($n$=sample size) rolling a dice using a random number table or generator or making $n$ draws from a hat/pile!
- (c) select units whose unit number corresponds to one of the $n$ random numbers.

---

**Exercise: Simple Random Sampling**

a random sample of 4 students to tell a fun fact about themselves

probability of selection?
Probability Sampling: How We Do It

In-class activity on Tuesday, 2/23:
- sampling from a telephone book

Probability Sampling: How We Do It

simple random sampling: how about large sample size?
- using a random sample function in statistical package.

Probability Sampling: How We Do It

stratified random sampling
- steps: (a) identify sub-populations (strata)
  (b) draw a random sample of each sub-population
- proportional stratified random sampling
  the proportion of each stratum = that in the population
- disproportional stratified random sampling
  the proportion of each stratum ≠ that in the population

Why Probability Sampling

- to ensure generalizability, i.e., the ability to generalize the conclusion based on the sample to the population
Non-Probability Sampling

availability sampling
select units that are available or easy to find.

snowball sampling
begin with one member of the target population, speaks to him/her, then asks that person to identify others.

Worksheet

Imagine you want to sample 100 members of the Freemasonry organization to interview them about something. Which of the sampling procedures described above might be most effective and efficient? Why?

Imagine you want to compare current Delaware residents who were born in the United States to current Delaware residents who were born in Peru. Which of the sampling procedures described above might be most effective and efficient? Why?