

- Collecting Data -

We consider 2 types of 'studies' in AP Statistics.

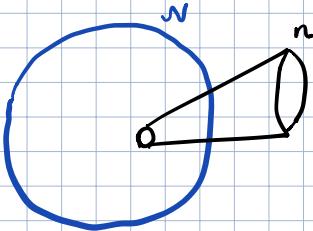
1) Observational Study: take a sample that is (hopefully) representative of a population.

2) Experiment: trying to show causal relationships between variables (more later)

Observational Studies

In an observational study we would like to take a sample of n observations from a population of size N , to make inference about the population.

- we would like our sample to be representative of our population.



Ex: Take a bite of an apple with your eyes closed. You don't need to eat the whole apple to know the fruit.

We have several techniques of sampling that we would like to explore. Let's look at all techniques through 2 lenses.

① - A scientist would like to know the proportion of voters in Alberta that vote for a specific party.

② - A scientist would like to know the average lifespan of Australian Kangaroos.

① Simple Random Sample:

random selection
random generator
random # table.

Select n observations at random from N . Note every observation and every sample must have equal likelihood of happening.

population of N
things and want to

sample n .

suppose $N=300$

$n=20$

r.#.t

51246 98121 45678

Σx

1. list elements of population

1. \rightarrow

2. \rightarrow

⋮

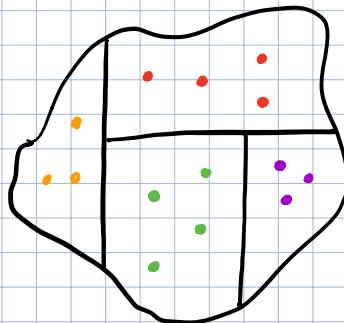
300. \rightarrow

1 - from a list of N voters,
randomly select n .

2 - Suppose that all kangaroos have a
tag with i.d. number. Put all i.d.
#s in a list and randomly select
 n . (Is this possible?)

② Stratified random sample.

Population has naturally occurring homogeneous
groups or 'strata'. Divide into strata, then randomly
sample from within.

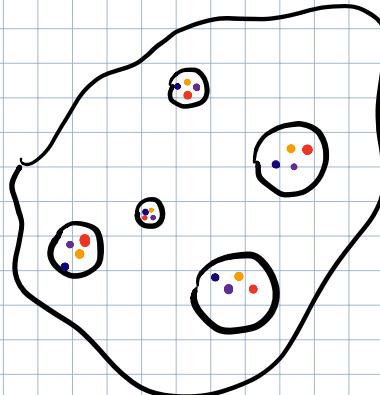


Σx 1 Divide population by
registered voting status.

2 Divide kangaroos into
male and female.

③ Cluster random Sampling

Population has naturally occurring heterogeneous clusters.
Randomly select some clusters and sample all points from
within.



Σx 1 Several towns with
heterogeneous make-up and sample
everyone.

2 several regions of
Australia are chosen as
clusters, all kangaroos are
sampled from within.

④ Systematic Sampling:

randomly select one observation from population, then sample every K^{th} observation.

Ex

- 1 from list of N voters randomly select one and then every K^{th} after.

⑤ Census

Sample entire population.

- 2 Need unique tags and then follow same procedure.



Non Random (Bad) Sampling

Convenience Sampling

Exactly how it sounds. Sample observations that are easiest to obtain.

Ex: 1 Pick up a phonebook and call n people for sample.

2 No worries, just let the Kangaroos come to you.

Voluntary Sample

People volunteer to be in sample. (often angry 'boomers')

Ex 1 Person puts paper ad asking about political ideas.

2 Kangaroos find you?

What is the best? Often SRS is the best choice, but can be costly or hard to obtain so we have other techniques.



Sampling Gone Wrong

- Sometimes our sample is not representative of the underlying population this is called bias.
- Non-random Sampling causes bias

Types of Bias → "undercoverage"

• Sampling (selection) Bias: Some types of observations have a higher likelihood of being sampled.

Ex ① - Researchers only sample from a list of NDP voters.

② - Just sample 'easy-to-obtain' Kangaroos. which may differ from other Kangaroos.

• Non-Response Bias: Observations are simply missed or not recorded.

Ex : ① - mail 5000 surveys out to respondents, but only 3500 return.

② - lose track of Kangaroos.

- Response Bias : Observations are simply wrong/False.
Can happen when a respondent lies.

Ex: 1 - People lie about who they will vote for

2 - Kangaroos lifespan is not recorded properly.

• Note: Some bias is simply out of a researcher's control. The aim is to minimize bias.
