# 4.1 Sampling & Surveys

A __population__ in a statistical study is the entire group of individuals we want information about. A __census__ collects data from every individual in a population. A __sample__ is a subset of individuals from a population from which we can collect data.

Ex: The student government at a High School surveys 100 students to get their opinions on a new bell schedule. What is the population and sample?

Sample: 100 high school students
population: All students at the high school.

How to sample ~~badly~~ poorly

① Convenience Sample: choosing individuals from a population who are easy to reach.

② Voluntary response Sample: consists of people who choose themselves to be in the sample by general invitation.
- typically attracts people who have a strong opinion for or against a topic.

The design of a statistical study shows <u>bias</u> if it would consistently under or overestimate the value you want to know.

* make sure you state the problem and classify it as an over/under estimate.

Good Sampling Techniques

① <u>Simple Random Sample (SRS)</u> of size $n$

is chosen in such a way that every group of $n$ individuals in the population has an equal chance to be selected as the sample.

Ex: Choose an SRS of size 4 from this class. start on line (0)

| | | |
|---|---|---|
| 00 Mitch | 06 Med. | 12 Jak. |
| 01 Abby D | 07 Lyd | 13 Sal. |
| 02 Abby H | 08 Car. | 14 Carm |
| 03 Fav. | 09 Br. | 15 Grac |
| 04 Rach | 10 Sam | |
| 05 Kir. | 11 E | |

Stratified Random Sample : classify the population into groups of similar individuals, called strata. Then we choose a separate SRS from each strata and combine these SRS's to form the sample.

- "same within, different between"
- stratified random samples give more precise estimates than SRS' of the same site
- make sure your strata make sense

<u>Cluster Sample</u>: classify the population into groups of individuals that are located near each other, called clusters. Then choose an SRS of the clusters. All individuals in the chosen clusters are in the sample.

- " Different within, same between"
- more practical than stratified random samples because it saves time and money
- chosen for ease, so there is as much variability as the population

p. 223 check your understanding

① Impractical due to the size of the arena and the space between the fans.

② Stratified by row letter since that gives me individuals who pay the same price for tickets

③ A numbered section will have a variety of fans who paid different prices

## Inference for sampling

① we can infer information about a population from what we know about a sample.

② Random sampling avoids bias in choosing a sample

③ You will probably not get a sample that exactly matches the population, a sample is an estimate and results vary from sample to sample.

④ How can we trust sampling? Laws of Probability

⑤ Larger random samples give better information about the population than smaller samples.

# Problems with sample surveys

① Undercoverage : occurs when some members of a population can't be reached.

② Nonresponse : occurs when an individual chosen for a sample can't be contacted or refuses to participate.

③ Response bias : a systematic pattern of of inaccurate. answers.

④ wording of questions

check your understanding p.228
① undercoverage
② Nonresponse
③ Undercoverage