

Lesson 25: Cluster and Stratified Samples

Stratified random sampling

Divide the population into groups of similar individuals (strata) then select an SRS within each strata. Combine the SRSs from each strata to form your full sample.

Goals:

- Guarantee that the sample represents the population.
- Split the population into groups (strata) **with meaningful differences** so the data represents the population.
- ***Each strata is different, but the subjects in the strata are similar.***
- The number to be surveyed in each strata should be proportional to the size in the population.

Advantage: Can produce more exact information (especially in large populations) by taking advantage of the fact that individuals in the same strata are similar to one another.

Disadvantage: Not appropriate unless strata are easily defined.

Example: In order to get a better idea of what athletes thought about homecoming last year, the director divides all athletes into the teams they play for, and then selects a random sample from each sports team. His full sample consists of aggregating the random samples from each team.

	Simple Example Population: 10 Men, 10 Women	Pros	Cons
Simple Random Sample	Place the names of all 20 people in a hat, shake it up and randomly draw out 8 names without replacement	We have an unbiased SRS. If the sample size is large enough, stratification is unnecessary.	The SRS of size 8 could result in one gender being over-represented and the sample may be biased toward that gender.
Stratified Random Sample	Strata 1 = men Strata 2 = women Place the names of the 10 men in a hat and the names of the 10 women in a second hat. Randomly select 4 men from the first hat without replacement and 4 women from the second hat without replacement.	We have an unbiased and representative SRS.	If the population is large and spread-out, it may be difficult to contact the subjects for data collection.

Who Wrote That Song?

Using a stratified random sample

A radio station that broadcasts a piece of music owes a royalty to the composer. The organization of composers (called ASCAP) collects these royalties for all its members by charging stations a license fee for the right to play members' songs. ASCAP has four million songs in its catalog and collects \$435 million in fees each year. How should ASCAP distribute this income among its members? By sampling: ASCAP tapes about 60,000 hours from the 53 million hours of local radio programs across the country each year.

Radio stations are stratified by type of community (metropolitan, rural), geographic location (New England, Pacific, etc.), and the size of the license fee paid to ASCAP, which reflects the size of the audience. In all, there are 432 strata. Tapes are made at random hours of randomly selected members of each stratum. The tapes are reviewed by experts who can recognize almost every piece of music ever written, and the composers are then paid according to their popularity.⁷

Cluster Sampling (Multi-stage sampling)

Divide the population into sections (clusters) then randomly choose a few of those clusters, and select every member or an SRS of the clusters chosen.

Goal:

- Make data collection easier.
- Split the population into similar groups so that each group/cluster represents the population.
- **Each clusters is similar, but the subjects within the clusters are different from one another.**
- The number to be surveyed in each cluster is the same for all clusters.

Advantage – Don't need a list of entire population

Disadvantage – More variability between samples depending on how clusters are determined.

Example – A psychologist at the University of Pennsylvania collects a sample by first dividing up the students into their respective schools (Wharton, engineering, nursing, arts and sciences) then by the departments that their major is in, and then she selects a few departments at random and surveys every student within those chosen departments.

	Simple Example Population: 10,000 High School students who live in our county, which has 15 High Schools.	Pros	Cons
Simple Random Sample	Place the names of all students in a hat, shake it up and randomly draw out 100 students without replacement. Survey these students.	We have an unbiased SRS that is likely to be representative of the population.	Some schools may be over-represented and the results will be biased towards these schools/students.
Cluster Sampling	Step 1: Randomly select 4 high schools from the county high schools to be clusters A, B, C, and D. Step 2: Place the names of all students from school A in a hat, shake it up and randomly select 25 students without replacement to be surveyed. Repeat this process for school B, C, and D.	We have an unbiased SRS that is <u>likely</u> to be representative of the population AND the data can be collected in a reasonable manner.	If each cluster is not representative of the population then we could end up the data that does not reflect the population. If all schools selected are suburban and no urban school is selected, then the sample is biased.

Daily Data Collection

Our school has 200 fall sport athletes on 12 different teams.

In groups: Create one question on a topic of interest and decide how to survey a sample using a stratified sampling and a cluster sampling. Volunteers will be called upon to share their methods to the class.

The class will decide on the most interesting question and use it to perform a survey in class.

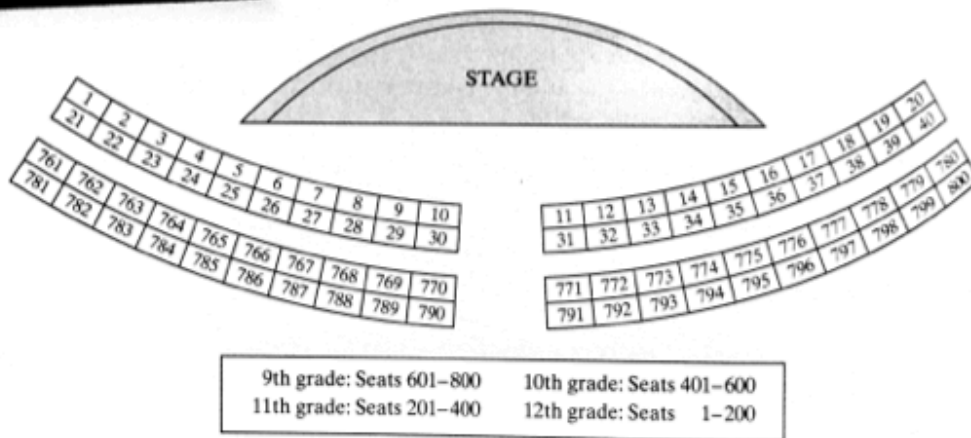
E Sampling at a School Assembly

Strata or clusters?



The student council wants to conduct a survey during the first five minutes of an all-school assembly in the auditorium about use of the school library. They would like to announce the results of the survey at the end of the assembly. The student council president asks your statistics class to help carry out the survey.

PROBLEM: There are 800 students present at the assembly. A map of the auditorium is shown below. Note that students are seated by grade level and that the seats are numbered from 1 to 800.



Describe how you would use each of the following sampling methods to select 80 students to complete the survey.

- Simple random sample
- Stratified random sample
- Cluster sample

CHECK YOUR UNDERSTANDING

The manager of a sports arena wants to learn more about the financial status of the people who are attending an NBA basketball game. He would like to give a survey to a representative sample of the more than 20,000 fans in attendance. Ticket prices for the game vary a great deal: seats near the court cost over \$100 each, while seats in the top rows of the arena cost \$25 each. The arena is divided into 30 numbered sections, from 101 to 130. Each section has rows of seats labeled with letters from A (nearest the court) to ZZ (top row of the arena).



- Explain why it might be difficult to give the survey to an SRS of 200 fans.
- Which would be a better way to take a stratified random sample of fans: using the lettered rows or the numbered sections as strata? Explain.
- Which would be a better way to take a cluster sample of fans: using the lettered rows or the numbered sections as clusters? Explain.