AP Statistics
Unit 03 – Sampling & Study Design
Types of Sampling

Name Key
Period ____

Random Sample: involves using a chance process to determine which members of a population are included in the sample.

Methods include:
- identical slips of paper in a hat with individual identifiers
- use a random # generator/table

Simple Random Sampling (SRS): sizen. Chosen in such a way that every group of n individuals in the population has an equal chance to be selected as the sample.

- gives equal individual change to be chosen
- unbiased
- every possible sample has an equal chance to be chosen

Choosing an SRS with TECHNOLOGY:

STEP 1 – LABEL: Give each individual in the population a distinct numerical label from 1 to N.

STEP 2 – RANDOMIZE: Use a random number generator to obtain n different integers from 1 to N.

You can also use a graphing calculator to choose an SRS.

TI-83/84

- Press MATH, then select PRB
- Choose/complete the command randInt(1, 1750)
- Press ENTER

Randomly generate 10 distinct numbers from 1 to 1750:

- Do randInt(1, 1750) again
- Keep pressing ENTER until you have chosen 10 different labels

If you have OS 2.55 or later, you can use the command RandIntNoRep(1, 1750) *to sort the numbers from 1 to 1750 in random order. The first 10 numbers listed give the labels of the chosen students.*

Choosing an SRS using Table B:

STEP 1 – LABEL: Give each individual in the population a numerical label with the same number of digits. Use as few digits as possible.

Example: We have 10 participants: label 0-9 instead of 01-10

Example: We have 100 participants: label 00-99 instead of 001-100

STEP 2 – RANDOMIZE: Read consecutive groups of digits of the appropriate length from left to right across a line in Table D. Ignore any group of digits that wasn't used as a label or that duplicates a label already in the sample. Stop when you have chosen n different labels.

Your sample contains the individuals whose labels you find.

Things to know about the table of random digits:

1. Each entry in the table is equally likely to be any of the 10 digits 0-9
2. Entries are independent of each other. Knowledge of one part of the table gives no information about any other part

Table B:

1. Groups and rows have no meaning
2. Label a possible random number between your highest and lowest numbers
3. Ignore numbers not in your label region
4. Enter table at any row/column unless told which row to start
5. Can read table B in any order: across, down, etc. (usually across)
6. Make sure each label had the same number of digits
7. Create the shortest possible labels
8. Start with 1, 01, 001, etc.

- label
- choose row
- choose direction
- choose how many, nonrepeating?

Example for how to use table B:

Joan's small accounting firm serves 30 business clients. Joan wants to interview a sample of 5 clients in detail to find ways to improve client satisfaction. To avoid bias, she chooses an SRS of size 5.

STEP 1 – LABEL: Give each client a numerical label, using as few digits as possible. Two digits are needed to label 30 clients, so we use labels 01-30.

It is also correct to use labels 00-29 or even another choice of 30 two-digit labels. Here is a list of the clients, with labels attached:

01	A-1 Plumbing	16	JL Records
02	Accent Printing	17	Johnson Commodities
03	Action Sport Shop	18	Keiser Construction
04	Anderson Construction	19	Liu's Chinese Restaurant
05	Bailey Truck	20	Magic Tan
06	Balloons Inc.	21	Peerless Machine
07	Bennett Hardware	22	Photo Arts
08	Best's Camera Shop	23	River City Books
09	Blue Print Specialties	24	Riverside Tavern
10	Central Tree Service	25	Rustic Boutique
11	Classic Flowers	26	Satellite Services
12	Computer Answers	27	Scotch Wash
13	Darlene's Dolls	28	Sewer's Center
14	Fleisch Realty	29	Tire Specialties
15	Hernandez Electronics	30	Von's Video Store

STEP 2 – TABLE: Enter table B anywhere and read two-digit groups. Suppose we enter at line 130, which is:

69051 64817 87174 09517 84534 06489 87201 97245

The first 10 two-digit groups in this line are 69 05 16 48 17 40 95 17

We only want to find digits between 01 and 30 so we would have to continue to find 5 random groups: 05 16 17 20 19

So the random groups that we selected using table B are
05 = Bailey Truck
16 =JL Records
17 =Johnson Commodities
20 = Magic Tan
19 = Liu's Chinese Restaurant

Example: Now you try it!

Kelly owns 26 fast food restaurants. She wants to survey a group of customers at 6 of her restaurant locations. She decides to use an SRS. Explain how she would choose the 6 restaurants to sample.

1. Label: label the 26 restaurants from 01-26

2. Randomize:
   a. Start: on row 131
   b. Direction: move from left to right
   c. Stop: after choosing 6 non-repeating
   d. Requirements: 2-digit numbers.

Systematic Sampling: members are selected from a random starting point at a regular (fixed) interval.
ex: every fifth person to walk into the gym.

Stratified Random Sample: classify the population into similar (homogeneous) groups called strata and do an SRS within each strata. Combine to form a complete sample.

Strata: Similar (homogeneous) groups of individuals within a population.

- Choose strata based on facts known before the sample is taken
- Should include non-overlapping groups: geographical areas, age-groups, gender

Example: The state of Oregon could be separated into counties

Example: BHS students could be separated by grade level

Cluster Sample: classify the population into groups of individuals that are located near eachother (clusters). then, choose an SRS of the clusters (entire clusters are part of the sample) and combine.

Clusters:

- groups of individuals that are located near eachother
- often heterogeneous groups that represent the population proportionally

- Clusters are most helpful when the cluster is representative of the overall population
- Often used for convenience because clusters are groups of individuals near each other

Example: Stratified vs. Cluster Sampling

Boston cream pie consists of a layer of yellow cake, a layer of pastry crème, another cake layer, and then a chocolate frosting. Suppose you are a professional taster whose job is to check your company's pies for quality. You'd need to eat small samples of randomly selected pies, tasting all three components: the cake, the crème, and the frosting.

One approach is to cut a thin vertical slice out of the pie. Such a slice would be a lot like the entire pie, so by eating that slice, you'll learn about the whole pie. This vertical slice contains all the different ingredients in the pie would be a *cluster sample*.

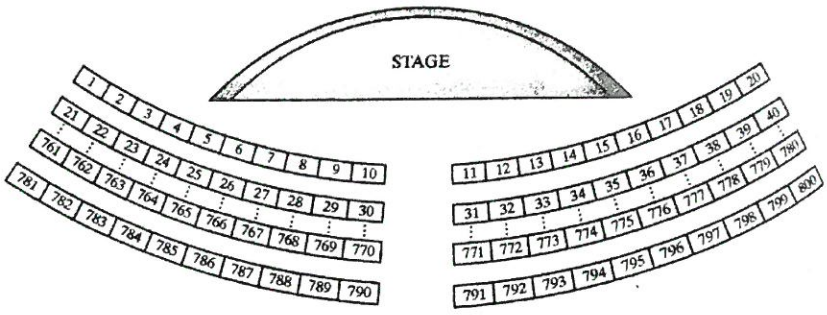
Another approach is to sample in *strata*; select some tastes of the cake at random, some tastes of the crème at random, and some bits of frosting at random. You'll end up with a reliable judgment of the pie's quality.

Many populations you might want to learn about are like this Boston cream pie. You can think of the subpopulations of interest as horizontal strata, like the layers of pie. Cluster samples slice vertically across the layers to obtain clusters, each of which is representative of the entire population. Stratified samples represent the population by drawing some from each layer, reducing variability in the results that could arise because of the differences among the layers.

Example: Sampling at a School Assembly

Strata or clusters? The student council wants to conduct a survey during the first give 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.

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



9th grade: Seats 601–800	10th grade: Seats 401–600
11th grade: Seats 201–400	12th grade: Seats 1–200

Describe how you would use your calculator to select 80 students to complete the survey with each of the following:

1) Simple random sample To take an SRS, we need to choose 80 seat #s at random. RandIntNoRep(1,800) and choose first 80. give survey to the students in those seats.
2) Stratified random sample
Strata = grade levels. let's choose 20 seat per grade. RandIntNoRep(601,800) x20
(401,600) x20
(201,400) x20
(1,200) x20
and give surveys to the students in those seats.
3) Cluster sample
Columns back from stage are clusters (kids from each grade). there are 20 clusters with 40 seats each. Choose 2 clusters. RandIntNoRep(1, 20)
and give surveys to all students in the 2 clusters chosen.

Inference: The purpose of a sample is the give us information about a larger population. The process of drawing conclusions about a population on the basis of sample data is called **inference** because we *infer* information about the population from what we *know* about the sample.

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

Undercoverage: occurs when some members in a population cannot be chosen in a sample.

- This causes bias!

Example: A survey of households: miss homeless people, prison inmates, & students living in dorms
Example: Opinion polls by phone miss 7-8% of Americans that don't have house phones

Example:

What would happen if we did a survey about increasing funding for Medicare by calling randomly selected landlines? Because many younger people only use cellphones, this method will likely overrepresent older people. Our estimate of the proportion of people who favor more funding will likely be too high.

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

- This causes bias!

Some students misuse the term "voluntary response" to explain why certain individuals don't response in a sample survey. Their idea is that participation in the survey is option (voluntary) so anyone can refuse to take part. What the students are describing is nonresponse. Think about it this way: nonresponse can only occur after a sample has already been selected. In a voluntary response sample, every individual has opted to take part, so there won't be any nonresponse.

Response Bias: A systematic pattern of inaccurate answers in a survey. people lie, misremember, make up answers.

Wording of Questions: most important influence on the answers given to a sample survey. Confusing or leading questions lead to strong bias. Changes in wording can greatly affect a survey's out come.

Example:

An opinion poll conducted in 1192 for the American Jewish Committee asked, "Does it seem possible or does it seem impossible to you that the Nazi extermination of the Jews never happened?" When 22% of the sample said "possible," the news media wondered how so many Americans could be uncertain that the Holocaust happened. Then, a second poll asked the question in different words, "Does it seem possible to you that the Nazi extermination of the Jews never happened, or do you feel certain that it happened?" Only 1% of the sample said "possible." The complicated working of the first question may have confused many respondents.