Unit 7, Lesson 3

Randomness in Groups

Let's explore why randomness is important in studies.

3.1 Study Selection

A reporter wants to know how people feel about the governor of her state. She decides to ask 100 people their opinions and thinks of several ways to ask the 100 people. For each method, explain the benefits and drawbacks, then choose the method for selecting 100 people that would best represent the people of the state.

1. Go to the capital city, and find 100 people interested in politics to respond to the survey.

2. Ask the 100 most politically influential people in the state to respond to the survey.

3. Obtain census data for the state and select, using a random process, 100 people from the list to survey.

4. Ask 50 registered voters who voted for the governor and 50 registered voters who did not vote for the governor to respond to the survey.

3.2 Hip Hop Memory

A research group interested in comparing the effect of different types of music on short-term memory gathers 200 volunteers for a study. One group will listen to a hip-hop music playlist while trying to memorize a list of 20 words. A second group will listen to a playlist of orchestral music while trying to memorize the same list of 20 words. After a break, the number of words recalled correctly by each individual is measured, and the results for the two groups are compared.

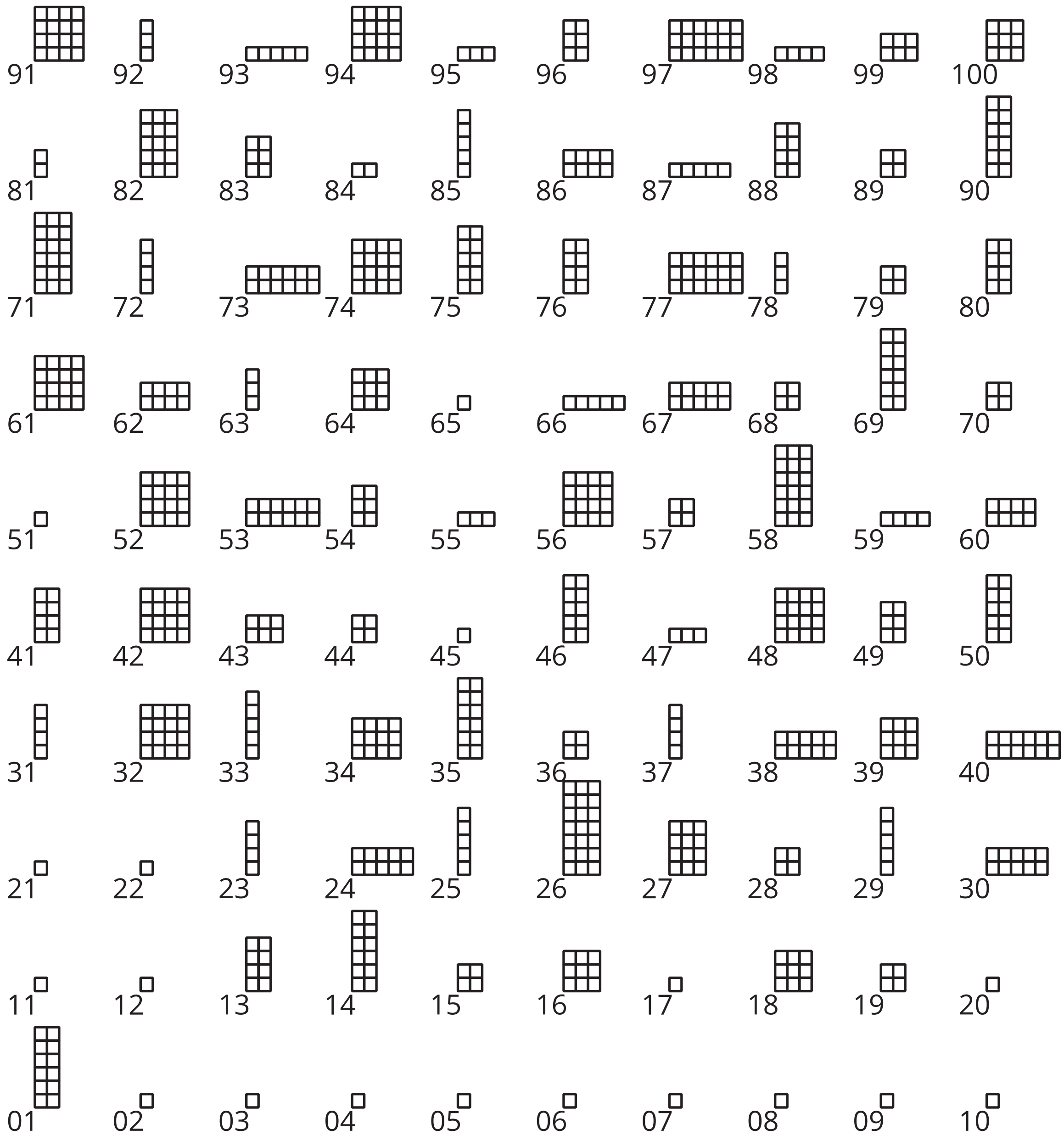
1. Is this an experimental study or an observational study? Explain your reasoning.

2. Which group do you hypothesize will recall more words? Explain your reasoning.

3. Here are some options for splitting the volunteers into groups. Which method will best address the intention of the study? Explain your reasoning.
   a. Divide groups based on volunteers' preferred music styles.
   b. Divide groups based on volunteers' ages. The youngest 100 people listen to hip-hop music, and the oldest 100 people listen to orchestral music.
   c. Divide groups based on the order in which volunteers come in to do the study. The first 100 listen to hip-hop music, and the next 100 listen to orchestral music.
   d. Write all the volunteers' names on slips of paper, put them in a jar and shake it, then draw out 100 slips. These volunteers will listen to the hip-hop playlist, while the others will listen to orchestral music.

3.3 Random Rectangles

A company offers solar power systems made up of 1-square-meter cells arranged into rectangles. They use the designs for their first 100 customers to list the ways people arrange the cells. They are interested in investigating this question: “What is the mean area of the rectangles created by our customers?”



1. Collect a sample of 5 rectangles using the methods here.
   a. Look briefly at the chart, and select 5 rectangles by their numbers. Record the numbers of the rectangles you choose.

   b. Select a number between 1 and 95. Use that number and the next 4 numbers for another sample of 5 rectangles. For example, if you select 8, then you would use rectangles 8, 9, 10, 11, and 12.

   c. Look closer at the rectangles, and choose your 5 favorites. Record the numbers of the rectangles you choose.

   d. Use a random number generator to select 5 numbers between 1 and 100.

2. For each method, find the mean area of the rectangles in the sample.

3. Which method do you think is best for estimating the mean area for the entire population? Explain your reasoning.

Are you ready for more?

How does a computer that runs predetermined instructions actually generate a “random number”? One way would be to try to connect the computer to something in nature that we consider random (like a number cube roll). This is doable, but generally not efficient, and the results cannot be replicated, so many computer programs use what is called "pseudo-random number generation." Essentially, they create lists of numbers that “seem” random, and for many purposes, that is sufficient.

Here is a version of one such method. Start with some number s (called the seed). To get the next number on the list, multiply the previous number by 6. If our new number is greater than 13, then divide by 13 and take the remainder. For example, if $s = 1$, our list of numbers is $1, 6, 10, 8, 9, 2, 12, 7, 3, 5, 4, 11, 1, \ldots$ Once we get back to our seed, our list will repeat.

1. What would our list be if we start with the seed $s = 2$? How does this relate to the list we had with $s = 1$?

2. What would our list be if we started with the seed $s = 1$, but instead of multiplying by 6 each time, we multiplied by 7 each time? Why does this list not seem as “random”?

Lesson 3 Summary

To get information about a characteristic of a population, people often measure that characteristic on a sample of individuals chosen from a population of interest. The idea is to draw conclusions about the population based on data collected from only the *sample*, a group that is selected from the whole population. To correctly generalize from the sample to the population, the researcher needs to know that the sample is *representative* of the population as a whole.

Suppose the engineer counts the pretzels only in the last 25 bags of trail mix that are produced on a particular day, and she finds that they contain too many pretzels. Should the engineer conclude that all the bags of trail mix produced that day contain too many pretzels? Not necessarily. Something might have happened late in the day that affected the number of pretzels in the bags. The last 25 bags of trail mix may not be a representative sample of the population.

So how do we get a representative sample? The best way is to let chance select the sample. For example, you might randomly select 25 different times throughout the day to remove the next bag of trail mix from the conveyor belt and count its pretzels. Using a process based on chance, in which each individual in the population is equally likely to be selected, is called **random selection** of the sample.

In experimental studies, it is often necessary to assign the individual participants in the sample to one or more groups. It is also best to assign individuals to groups using a random process.

For example, say that you were studying the effect of students turning off electronic devices while doing homework. After a representative sample is selected, you need to assign the individuals in the sample to two groups: one group makes no changes to the conditions in which they normally do homework, and another group that turns off electronic devices while doing homework. Examples of assignment processes that are *not* random include:

- Assigning students whose names start with A–L to one group and M–Z to the other group
- Assigning students who play a musical instrument to one group and the rest to the other group
- Asking for volunteers to be part of the group that turns off electronic devices

In order to assign individuals randomly to groups, every individual must have an equal chance of being assigned to either group. Examples of assignment processes that *are* random include:

- Writing each participant's name on a slip of paper and mixing the slips well in a bag. Drawing half of the names from the bag and assigning these participants to one group, and the rest to the other group.
- Flipping a coin for each participant, and placing them in one group if the result is heads and the other group if the result is tails
- Writing a list of participants and numbering the list, then a random number generator to select participants for one group

When subjects are not assigned to experimental groups using a random process, other factors may influence the results from the experimental study so that the data does not answer the initial question. In this example, if the groups are split by volunteering, the impact of turning off the devices may be impacted by similar traits of the subjects who volunteer, such as their not using electronic devices much already or having a personality that is willing to volunteer to try something new. These traits may influence the results so the data from the experimental study do not accurately address the question about the impact of electronic devices on student homework.