Why Do We Sample?

Estimated Time: 30-40 minutes

SUMMARY

An average is a single number that represents a list of numbers or measurements. Although there are several ways to represent a list of numbers or measurements with a single number, an average generally is the mean. A mean is calculated by the sum of all numbers in the list divided by how many numbers are being averaged. In many cases, the entire list of numbers is used to calculate an average. For example, if your teacher calculates the average for the class on the test, your teacher knows how everyone in the class did. But, how do we take averages for really long lists of numbers? For example, how do we know the average height of all 12 year old kids in the United States? There are a lot of 12 year old kids in the United States. It would be impossible to count all of them! In a case like this, we would need to sample the population. But, how do we know how to sample? When is our sample large enough to represent the population?

WHAT YOU'LL LEARN

- The importance of sampling a large population.
- How sampling can affect what we know about a population.

Materials Used	**Resources Used**
• Bag of M&Ms or other candies of various colors • Bowls to sort candies • Calculator	• The Importance of Sampling: https://www.youtube.com/watch?v=Gs-gLeYuDZw

WHAT TO DO

1. Sort and count the number of each candy color you have. I used a 1-pound bag of M&Ms, so that is the example I will be using throughout this activity. Table 1 has all the data for my 1-pound bag.

Table 1. Counts and percentages of candy colors in 1-pound bag of M&Ms.

Brown	Brown	Blue	Orange	Yellow	Green	Red	Total
Count	50	140	139	92	115	86	622
Percent	8.0%	22.5%	22.3%	14.8%	18.5%	13.8%	100%

2. The total number of candies represents your entire population. It probably took a pretty long time to sort and count. What if you want to try to find what the population looks like with a sample. Let's try that out!
3. After counting and sorting your candies (and recording these values), recombine your candies and mix them all up. Now, randomly pull out 10 candies.

Table 2. Counts and percentages of candy colors of 10 count candy samples.

Sample	Brown	Blue	Orange	Yellow	Green	Red	Total
1 (count - %)	1 - 10%	2 – 20%	2 – 20%	1 – 10%	2 – 20%	2 – 20%	10 – 100%
2 (count - %)	0 – 0%	2 – 20%	3 – 30%	3 – 30%	2 – 20%	0 – 0%	10 – 100%
3 (count - %)	1 – 10%	2 – 20%	1 – 10%	1 – 10%	1 – 10%	4 – 40%	10 - 100%

4. Compare your sample of 10 candies to your population. How does each sample compare? What is over represented and what is underrepresented? How might you make your sample look more like the actual population?
5. Recombine your samples of 10 back into your entire bag. Now, let's see what happens if we sample 3 groups of 20 candies out of the population.

Table 3. Counts and percentages of candy colors of 20 count candy samples.

Sample	Brown	Blue	Orange	Yellow	Green	Red	Total
1 (count - %)	2 - 10%	4 – 20%	7 – 35%	3 – 15%	3 – 15%	1 – 5%	20 – 100%
2 (count - %)	1 – 5%	3 – 15%	4 – 20%	3 – 15%	6 – 30%	3 – 15%	20 – 100%
3 (count - %)	2 – 10%	6 – 30%	5 – 25%	1 – 5%	4 – 20%	2 – 10%	20 - 100%

6. Compare your sample of 20 candies to your population. How does each sample compare and how does the samples of 20 compare to the samples of 10? What is over represented and what is underrepresented? Do the samples of 20 look more like your actual population?
7. You can repeat sampling the candies to see what sample size you would need to percentages that are representative of your actual population.

TIPS

- Another interesting thing that happens in biology can also be demonstrated with a bag of M&Ms. The founder's effect is a mechanism of evolution where a small number of the population is cut off of the main population. As the small number is cut off, they have a limited number of genes in that small population to be passed onto the next generation. You can demonstrate founder's effect by sampling a small number of M&Ms from your bag, say about 20. Then, have the small population replicate. In this case, you would add the same number of browns, blues, oranges, yellows, greens, and reds to the next generation as was in your original small sample. Repeat this process a few more times. Then, compare your new population that is a multi-generational replication of your original small population to the population in the bag of M&Ms. Did any of the colors (genes) fall out of the population? Does the new population have an over or underrepresentation of particular colors?