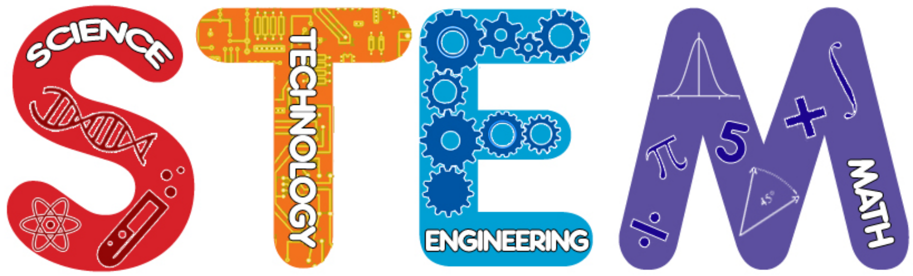
The Beadery® Bead Mix Statistics



Materials Needed:

- Your favorite mix of multicolored beads

Tools Needed:

- Pen / Pencil
- Piece of paper
- Scoop (can be a large spoon, measuring cup or even your hand)
- Calculator

Skills Learned:

- Counting/Data Collection
- Introductory Statistics

Background:

Statistics uses numerical information and data collection to describe properties of a group, called the p*opulation*. This data collected often shows trends and patterns that can be used to make models of a system or make predictions of future outcomes.

Precedure:

1. Create a *frequency table* on your paper with columns labeled "***Color***", "***Sample 1***", "***Sample 2***", "***Sample 3***" and "***Total***". Leave enough room under each header to have a ***"Count"*** and ***"%"*** column.
2. List all of the colors in your bead mix under the "***Color***" column.
3. Add a row to the bottom of the table to record the "***Totals***".

Example Frequency Table (6 Bead Colors):

	Sample 1		Sample 2		Sample 3		Total	
Color	**Count**	**%**	**Count**	**%**	**Count**	**%**	**Count**	**%**
Red								
Orange								
Yellow								
Green								
Blue								
Purple								
Total								

4. Take a scoop of beads and place them on a tray or work surface that will prevent rolling. You can use anything you like for a scoop. Measuring cups or ice cream scoops work well depending on the size of the beads or you can even use your hand. Each scoop doesn't need to contain exactly the same number of beads, just enough to get a good *sample* of colors.
5. Count the number of each color bead in the sample and record it in the "***Count***" column for that color. This is called the *frequency* of the color for that sample.
6. After you have counted each of the colors in the sample, add up the total number of beads in the sample and record it in the "***Total***" box. This is your *sample size*.
7. Calculate the percentage of the mix each color makes up and record it in the "***%***" column. The percentage, also called the relative frequency or *probability*, is calculated by dividing the count of each color by the total count in the sample.

$$Color\ Percentage = \left(\frac{Color\ Count}{Total\ Count}\right) \times 100$$

   For example, if your mix has 12 red beads in a sample of 77 beads:

$$Red\ Percentage = \left(\frac{12\ red\ beads}{77\ beads}\right) \times 100 = 15.6\%$$

8. Repeat the data collection for the remaining samples.
9. Calculate the probabilities for all the bead samples taken by adding up the counts of each color sample and putting them in the "***Total Count***" column. Use these frequencies to calculate the probabilities in the "***Total %***" column.

Observations and Results:

- *How do the color frequencies of each of the samples compare to the color frequencies of the total population?*

When we make our bead mixes at The Beadery, we generally use equal amounts of each color in the mix. Therefore, an ideal mix of six colors should give you a uniform distribution of $\frac{1}{6}$ (approximately 16.6%) of each color. However, the process has many factors that may result in variations in these ratios in your sample. To make a typical bead mix, we take millions of beads at a time, mix them in a giant tumbler and then package them in smaller quantities. Each bag of beads, which is your population, will probably not end up with a perfectly equal mix. This might lead to your sample data showing one or two colors that show up with a higher frequency than others, so don't worry if your frequencies are not equal. Finding out the actual color makeup of an unknown batch is one reason why you would do a statistical analysis in the first place.

- *What do you think would happen to the accuracy of your sample frequency if you counted fewer/more samples?*

The more samples you take, the closer you get to counting the complete population. Therefore, the higher the number of samples, the higher the accuracy of your total frequency calculations. If you take fewer samples, your data will be less representative of the total population and the accuracy might go down.

- *What would happen to the accuracy of your sample frequency if each sample were smaller/larger?*

Similar to increasing the number of samples taken, increasing the sample size will get you closer to counting the complete population quicker. This may give you a more accurate representation of the frequency with fewer samples. Smaller samples might not have enough beads to represent the total distribution of colors, so the accuracy will be lower per sample and you would need more samples to get a better representation of the population.

- *How many samples will you need to take before you can acceptably predict the probability of each color in the total population or the next sample?*

The first thing you must remember is that due to randomness, you will almost never get a "perfect" sample that represents the total population exactly without counting the entire population. You will have to decide how much of a margin of error is acceptable in your situation. The more confidence you need in the data, the more samples you will need to take.

How many samples you need to take will depend on several factors, such as the size of each sample, the size of the total population, the number of different colors, etc. A simple experiment that you can do to estimate this is as follows:

1. Set up a control mix of beads that you know the exact frequency of each color. This mix can be all equal quantities of each color or any distribution that you want.
2. Take a sample of the mix. Count the frequency and calculate the probability of each color.
3. Take a second sample, count the beads and calculate the probabilities of each color in that sample and the averages for both samples taken.
4. Repeat with more samples until you have the total average probabilities that are within your acceptable range of confidence. If you are comfortable with a *confidence level* of ±10%, you will reach that with fewer samples than if you want only ±5%.

5. Create another control mix of beads with a different known color distribution.
6. Take the number of samples that you decided were "acceptable" in Step 4, recording the data in a frequency table and calculating the probability of each color.
7. Calculate the average probability of each color for all of the samples taken. How do these values compare to the known distributions of the mix? Are you within your acceptable confidence level?

Vocabulary:

Confidence Level – the probability that the true, unknown characteristic is within your acceptable range

Data – observations collected from an experiment or study

Distribution – the probabilities of all possible outcomes of an experiment, usually represent as a graph

Frequency – the number of times a particular value or characteristic occurs in a sample

Frequency Table – a table of the number of data points in a sample that have a certain value or characteristic

Population – the collection of all objects from which data can be selected

Probability – the likelihood that a particular event will occur, usually expressed as a percentage between 0% (impossibility) and 100% (certainty)

Sample – a subset of the total population

Sample Size – the number of items in a sample

Statistics – the branch of math that deals with the collection, organization and interpretation of data