West Virginia math4life

Domain	Probability and Informed Decisions	
Cluster	**Use probability to evaluate outcomes and make decisions.**	
Standard(s)	M.ASHS.22	Connect sampling variability and margin of error to generate and interpret plausible parameter values. Instructional Note: The concept of statistical significance is developed informally through simulation as meaning a result that is unlikely to have occurred by chance alone. Focus on statistics as a way of dealing with, not eliminating, inherent randomness.

Content Examples:

- Introduction to Sampling Distributions: *https://www.youtube.com/watch?v=z0Ry_3_qhDw*
- Introduction to Confidence Intervals and Margin of Error: *https://www.youtube.com/watch?v=hlM7zdf7zwU*

Relevant Content:

Vocabulary:

- Cluster Sample: A cluster sample is a sample obtained by classifying the population into groups of individuals that are located near each other, called clusters, and then choosing a simple random sample from the clusters. All individuals in the chosen clusters are included in the sample.
- Confidence Interval: A confidence interval gives an interval of plausible values for a parameter. The interval is calculated from the data and has the form:

  point estimate ± margin of error,

  or alternatively,

  statistic ± (critical value)(standard deviation of statistic).

- Critical Value: A critical value is a multiplier that makes a confidence interval wide enough to have the stated capture rate. The critical value depends on both the confidence level, C, and the sampling distribution of the statistic.
- Margin of Error: The margin of error represents the difference between the point estimate and the true parameter value. This difference will be less than the margin of error in C% of all samples, where C is the confidence level.
- Population: A population is the entire group of individuals we want information about in a statistical study.

» Randomization Distribution: A randomization distribution is a distribution of a statistic in repeated random assignments of experimental units to treatment groups assuming that the specific treatment received does not affect individual responses.

» Sample: A sample is a subset of individuals in the population from which we collect data.

» Sampling Distribution: A sampling distribution is the distribution of values taken by a statistic in all possible samples of the same size from the same population.

» Simple Random Sample: A simple random sample is a sample chosen in such a way that every group of n individuals in the population has an equal chance to be selected as the sample.

» Statistic: A statistic is a number that describes some characteristic of a sample.

» Stratified Random Sample: A stratified random sample is a sample obtained by classifying the population into groups of similar individuals, called strata, then choosing a separate simple random sample in each stratum and combining these simple random samples to form the sample.

Assessment Links or Tasks:

» Investigating Confidence Interval:
http://www.apstatsmonkey.com/StatsMonkey/ReadBestPractice_files/Hersheys%20Kiss%20Activity.docx

» Sampling Distribution Applet:
http://www.rossmanchance.com/applets/OneProp/OneProp.htm?candy=1
See the following instructions: *Orange Candy Applet Instructions*
Student activity: *Orange Candy Student Activity*

» Sampling Distribution Applet 2: *http://onlinestatbook.com/stat_sim/sampling_dist/*

» Random Babies Applet: *www.rossmanchance.com/applets/randomBabies/RandomBabies.html*
See the following instructions: *Random Babies Instructions*

» One Variable Sampling Applet:
http://www.rossmanchance.com/applets/OneSample.html?population=gettysburg
See the following instructions: *One Variable Sampling Instructions*

Orange Candy Applet Overview

In this activity, we will create a sampling distribution of the sample proportion of orange candy pieces. Students may choose:

(1) the size of the sample;
(2) the number of samples;
(3) the number of orange candies (or the proportion of orange candies);
(4) to animate the selection process (or not animate the selection process);
(5) to display summary statistics; and
(6) hide (or display) the candy dispenser.

This applet will provide students will a visual representation of a sampling distribution of a sample proportion. Students will notice that the mean of the sampling distribution is the value of the population parameter (the true population proportion of orange candy pieces).

Students will also discover that changing the size of the sample will ultimately change the shape of the sampling distribution, but not the center. As the size of the sample increase, the standard deviation of the sampling distribution decreases.

Orange Candies Lab

West Virginia math4life

NAME

Preliminary Questions:

P1. If you take a sample from the population of all orange candies, would you expect your sample to have the same proportion of orange pieces as the population of all orange candies? Explain.

P2. How close would you expect your sample proportion ($\hat{p}$) of orange candies to be to the true proportion (p) of orange candies? Answer using standard deviations.

Will it always be this close? Explain.

P3. Assume you randomly select 25 orange candies and find that 12 are orange. That means that 48% of your sample were orange. How far away would you expect the true proportion of orange pieces to be from 48%? (Answer again in standard deviations.)

Will it always be this close? Explain.

P4. How far away in percents would you expect the true proportion to be? First, we need to calculate the standard deviation "in percents." What is the formula for the standard deviation of a sample proportion?

$$\sigma_{\hat{p}} = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_$$

- But we don't know what ____ is, do we? So instead, we will use the next best thing, _____.
- So for our example, we expect the true proportion of orange pieces to be within two standard deviations or _______ from my sample proportion of .48.
- Thus, we are "pretty sure" that the true proportion lies somewhere between _______ and _______.
- This interval is called a ***confidence interval***.

P5. How sure are we? ______ Therefore, this is called a ______ confidence interval.

Let's get some data!!

1. From the tub of Orange candies, collect a small cupful (3oz) of pieces.

2. At your desk, randomly select two or three samples of 25 orange candies to analyze.

3. Record the number of pieces in each color in the tables below.

Samp 1	Org.	Brn.	Yel.
# Pieces			
Percent			

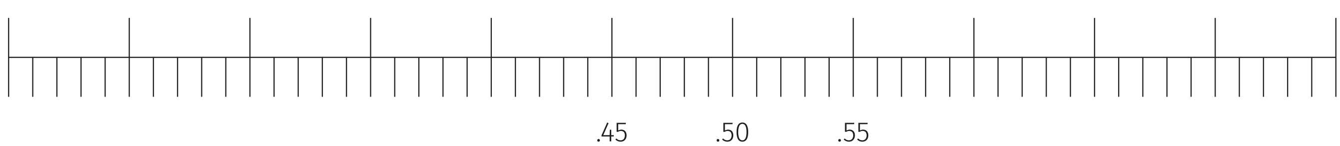
Samp 2	Org.	Brn.	Yel.
# Pieces			
Percent			

Samp 3	Org.	Brn.	Yel.
# Pieces			
Percent			

4. We are trying to find out what percent of Orange candies are orange. Let $\hat{p}$ be the proportion of orange candies in your sample. Fill out the table below, calculating $\hat{p}$, the standard deviation of $\hat{p}$, and a 95% confidence interval.

Sample	$\hat{p}$	SD of $\hat{p}$	95% Confidence Interval
1			
2			
3			

5. On the board, place a dot for each $\hat{p}$ you calculated. Record the class distribution below:



6. Describe the sampling distribution of $\hat{p}$.

7. Since this distribution is relatively ________________, we can expect our samples of 25 Orange candies to behave according to the characteristics of a ______________ distribution.

   And according to the Central Limit Theorem, the mean of our sampling distribution, which is about _______ should be (about) equal to the mean of the ________________________. (It would be exactly equal to it if we gathered all possible samples of 25.)

   So if our samples behave "normally," then we would expect to land within two standard deviations of the mean about ______% of the time.

   Or another way to say it: I'm pretty confident (95% confident to be exact) that the true proportion of orange candies is within two standard deviations of my $\hat{p}$.

   To be more specific, if your 95% confidence interval was (.28416, .67584), then you could say that you are 95% confident that the true proportion of orange candies is between .284 and .675.

8. To summarize, we can calculate a confidence interval by collecting a random sample, making sure the sample size is correct (more on that later), calculating $\hat{p}$, and adding/subtracting a certain number of standard deviations from $\hat{p}$.

   The number of standard deviations you add/subtract will be determined by your ___________________. For a 95% CI, you need to use ________ standard deviations. For 90% confidence, you need to use ________ standard deviations. These can always be found at the bottom of a t-table and are considered z-scores. So the general formula for confidence intervals for proportions is:

   _______ ± _____ · _______________

   There are three conditions that need to be met:

   1.

   2.

   3.

9. The number you add/subtract from $\hat{p}$ is called the _______________ _____ ______________.

_______ ± _____ · _______________ ←

So the margin of error formula is: ________ · _______________

10. The MOE formula is helpful if we need to calculate a certain sample size that would be needed to keep the MOE under a certain amount. What were the margins of error for your CI's? Many were over .10, right? In fact, you might say that our margins of error were pretty large—not very accurate for predicting the true percent of orange candies.

    So let's assume that we wanted to reduce our MOE in the orange candies problems to .05. That would mean that we would write an equation setting MOE ≤ .05

    If my $\hat{p}$ was .48, then my equation would be:

    $1.96\sqrt{\frac{.48(1-.48)}{n}} \leq .05$ Solve this equation for n!

n ≈ 383.5 or since it is a sample size, we would say 384 (ALWAYS round up).

Good news! We'll need more candies in our samples!

Random Babies Applet Overview

In this activity, we will create a bar chart and a plot of the average number of correct matches. Students may choose:

(1) the number of babies;
(2) the number of trials;
(3) to display the theoretical probabilities (or not); and
(4) to animate the selection process (or not animate the selection process).

This applet will provide students with a visual representation of a randomization distribution.

One Variable Sampling Applet Overview

In this activity, students will compare the shape, center, and spread of a population model with a sampling distribution. Students may choose from the following data sets:

(1) Gettysburg Address word length;
(2) age of pennies;
(3) star density; and
(4) Beyonce's Crazy in Love word length.

Students may also choose:

(1) stratified random sampling;
(2) cluster random sampling;
(3) the number of samples; and
(4) the number of clusters.

This applet will provide students with a visual representation of a randomization distribution.