SRS vs. Convenience Sample in the Gettysburg Address

Written by: Jim Spiker and Trey Cox

Arizona Mathematics Partnership

[James.spiker@cgc.edu](mailto:James.spiker@cgc.edu)

# Overview of lesson

Students have an interesting view of what a random sample looks like. They often feel that just closing their eyes and picking “haphazardly” will be enough to achieve randomness. This lesson should remove this misconception. Students will be allowed to pick words with their personal definition of random and then forced to pick a true simple random sample and compare the results.

# CCSS

CCSS.Math.Content.6.SP.A.1

CCSS.Math.Content.6.SP.B.4

CCSS.Math.Content.6.SP.B.5

CCSS.Math.Content.7.SP.A.1

CCSS.Math.Content.7.SP.A.2

Almost all of the CCSS are presented in this lesson

# Prerequisites

Students should be able to construct dot plots, box and whisker plots, find averages, and find the mean absolute deviation (MAD).

# Learning Targets

Students should learn the value of taking a SRS over a convenience sample. Students should see the effect of taking larger samples on the precision of our estimate for the mean of the population.

# Time Required

Approximately one class hour should be allowed for this exercise especially if the extension on variation is performed.

# Material Required

The Gettysburg Address handout and a graphing calculator or other graphing technology could be used. Post-it notes. Two different colors would be best.

# Lesson Details

The Gettysburg Address is a famous oratory delivered by Abraham Lincoln. What is the readability of this document? Researchers often rate the readability of a document by finding the average word length of the passage. Since this document has over 250 words in it we will take a random sample of words and take the average of that sample to use it as an estimator of the mean of the entire passage.

Have the students look at the handout and ask them to pick 5 words at random. Do not prompt them on how to make this happen. Allow students to create a random sample by any method they choose. Allow them to discuss how they want to do it. Most will decide to simply close their eyes and pick 5 words.

After they have chosen their 5 words have them count the letters in each word and find the average of this sample. ROUND TO THE NEAREST TENTHS PLACE. Now have them plot each of their averages as a point on a dot plot or a histogram on the board (we will use post-it notes). Now have a discussion on what type of sample most of the students have chosen. If they simply close their eyes and stab at the document do they think they might have some bias towards smaller or larger words?

So with this in mind, discuss the definition of bias. Since they have methodically chosen larger over smaller words they will probably be overestimating the average word length and thus they will make the document seem to be more challenging to read. Lead them through a discussion about this.

Now have the students use technology like the random number generator on the TI 83/84 by entering math-prb-randint and enter randint(1,268,5) This will select 5 random numbers. The handout has the words in the address associated with each number from 1 to 268 and thus they have allowed the calculator to select the words for them. Also every word was equally likely to be chosen and thus we will not have any bias towards larger or smaller words. Therefore, we have more reliable data and can make a better decision.

Next have the students calculate the average of their sample and plot the sample on a new dot plot or histogram on the board. Compare this graph to the convenience sample graph. Have the students estimate what the average word length is from their first distribution and compare it to the second distribution. If all has gone well the SRS should have a mean that is closer to the actual mean which is 4.295 letters per word. Now you can ask the students what will happen if they take a SRS of size 10. What do they think will happen to their estimate of µ? Will they get a better or a worse estimate and why? What do they think will happen to the amount of variability in the sampling distribution? Will the amount of variation increase or decrease? Now go ahead and have the students pull the SRS of size 10 and plot the dot plot as a class. Compare the mean with the SRS of size 5 and see if the estimate is better. Now have the students construct the MAD for both the SRS of n=5 and n=10. Have them compare the size of the MAD. What do they think will happen if we sample with n=30?

# Student Handout

The Gettysburg Address is a famous oratory delivered by Abraham Lincoln. What is the readability of this document? Researchers often rate the readability of a document by finding the average word length of the passage. Since this document has over 250 words in it we will take a random sample of words and take the average of that sample to use it as an estimator of the mean of the entire passage.

***Lincoln’s Gettysburg Address***

1. Directions-Circle 5 words from the Gettysburg Address at random. Count the number of letters in each word and compute the mean number of letters per word. Put this value on the post-it note provided and then place this post-it note on the board to create a histogram.

*Four score and seven years ago our fathers brought forth upon this continent, a new nation, conceived in Liberty, and dedicated to the proposition that all men are created equal. Now we are engaged in a great civil war, testing whether that nation, or any nation so conceived and so dedicated, can long endure. We are met on a great battlefield of that war. We have come to dedicate a portion of that field, as a final resting place for those who here gave their lives that that nation might live. It is altogether fitting and proper that we should do this. But, in a larger sense, we cannot dedicate we cannot consecrate we cannot hallow this ground. The brave men, living and dead, who struggled here, have consecrated it, far above our poor power to add or detract. The world will little note, nor long remember what we say here, but it can never forget what they did here. It is for us the living, rather, to be dedicated here to the unfinished work which they who fought here have thus far so nobly advanced. It is rather for us to be here dedicated to the great task remaining before us that from these honored dead we take increased devotion to that cause for which they gave the last full measure of devotion that we here highly resolve that these dead shall not have died in vain that this nation, under God, shall have a new birth of freedom and that government of the people, by the people, for the people, shall not perish from the earth.*

1. Use the calculator to pick 5 random numbers from 1 to 268 by typing math-prb-randint(1,268,5) or use a random number table. Find the words associated with these numbers in the table attached and compute the average word length as you did in part 1. Write this value on the post-it note provided and put it on the histogram on the board.



1. Write a few sentences that compare and contrast the two dot plots. How are they alike and how are they different?
2. Now repeat the process of using the calculator but this time create a random sample of size 10 by typing math-prb-randint(1,268,10) or use a random number table. Find the average word length as you have before and make a new histogram on the board. What do you notice about the variation in the distribution? Has it increased or decreased? Why do you think that happened?
3. For each dot plot combine all of the classroom data into a list. Use each list to compute the mean, median, and MAD of each distribution. How are these values alike? How are they different? What do they each tell us?