

Introduction to Statistics I

Instructor: Jodin Morey moreyj@lemoyne.edu

Previous Lecture

- ◆ Sampling from a Population & Sampling Bias
- ◆ Explanatory/Response/Lurking/Compounding Vars
- ◆ Observational Studies & Cause/Effect



Topic 4: Random Sampling



RQ: What's the average length of a word in the Gettysburg Address?

Obv-unit/var/Population/Parameter?

Poll:



bit.ly/introstatsdata

Poll: Gettysburg Address

Activity: 4-1

Obv-unit: words in Address. **Variable:** word length.

Population: all words in Address. **Parameter:** average length of words in Address.

Sampling

Let's say we want to limit our sample to ten words from the Address. Statistic?

Statistic: average length of words in our ten word sample from Address.

Four score and seven years ago our fathers brought forth on 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 battle-field 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 can not dedicate—we can not consecrate—we can not 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.



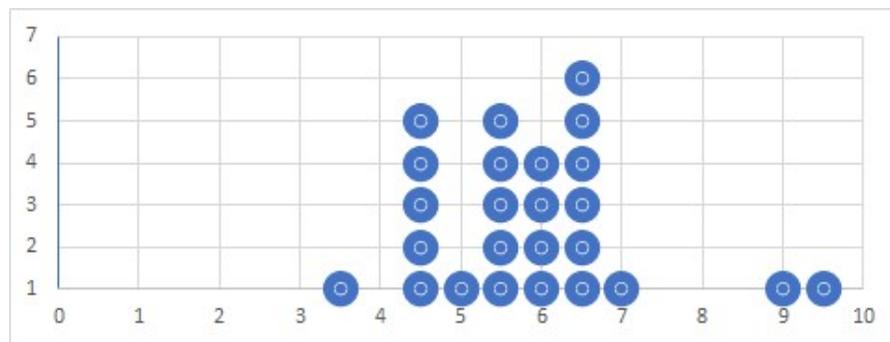
Each student: select 10 words from above address that appear to be of typical length.

Then, report average length of your ten words.

bit.ly/introstatsdata

Poll: Gettysburg Address

Average Word Length in 10 Word Samples



Our sampling method is **biased**. On average, we ended up with statistics that were consistently high (actual average is 4.29). We **overestimated** the parameter.

How might we sample better?

What if we closed our eyes and picked 10 words that way?

Alternative Sampling Method

Assign each word in the Address a number. (see table on pg58, and Table I in back of book)

Use a computer to sample ten random numbers; select those ten words.

(check "Show Sampling Options", then "Draw Samples")



bit.ly/introstatsdata

Applets: Sampling Words

Simple Random Sampling (SRS): sampling methodology in which all members of a population have an equal chance of being selected for a sample. SRS is an unbiased sampling method.

A sampling method is unbiased if statistics drawn from different random samples are centered at parameter.

Recall: What leads to biased samples?

◆ Sampling from only (nonrepresentative) part of the population



◆ Convenience Samples (lazy sampling)



◆ Voluntary Samples (self-selection)



◆ Allowing non-response



Most Straitforward SRS: Accomplished by numbering the members of the population using a **sampling frame**.
(often not available)

This activity also illustrates concept of **sampling variability**: the idea that the value of a statistic varies from sample to sample.

What did we learn?

- ◆ Simple random sampling (SRS),
- ◆ Sampling variability
- ◆ We want unbiased/consistent sampling method

