Introduction to Statistics
Activity 1.1 **SOLUTIONS**

Gettysburg Address

a) Pick ten words at "random" from the Gettysburg address on the screen. Record the word and how many letters it has below:
Responses will vary.

b) Indicate the observational units and variable.
Observational unit: **Words in the Address.** Variable: **# of letters.** Type: **Quantitative.**

c) What is the average (mean) number of letters in your sample? Is this a parameter or statistic?
Responses will vary. Statistic.

d) Combine the sample average with those of your classmates.
[Depends upon student choices that day]

e) Indicate the observational units and variable for the plotted data from part d.
What do the dots represent?
Observational unit: **Classmates.** Variable: **Mean # of letters in sample.**

f) The actual average letter length in the Gettysburg address is 4.29. Add that to your class average dotplot. How many students had a larger average? What's the proportion?
[Depends upon student choices that day]

g) Was the sampling method biased? In what direction? How can you tell?
Yes. Student estimates tend to be overestimates. One can tell this by noting that more of the averages are bigger than the actual average of 4.29.

h) Why is the sampling method biased?
Yes.

i) What about closing your eyes and pointing at words at random – would this be biased? If so, why?
Folks might tend to point toward the center of the paragraph, and avoid the edges. Also, larger words take up more room and therefore are more likely to be pointed at than smaller letters.

Our selection method should give every member of the population an <u>equal chance</u> of being in the sample. This is called <u>simple random sampling</u> (SRS)

j) We are going to use a computer-generated table of random digits to determine which words to include in our sample. Go to random.org and ask the tool to pick numbers between 1 and 268. Generate 5 numbers and put them in the "SRS #" column. Then, use the table on the overhead to find the associated word, then determine its length.
Responses will vary.

k) What is the average length of words in your sample?
Responses will vary.

l) Combine the sample mean with those of your classmates.
Responses will vary.

m) How does the distribution of "chosen" sample averages compare to "random" sample averages?
You will find that the chosen averages are greater than the random averages. This is due to the biases identified above.