STATISTICS TEACHER
SUPPORTING THE TEACHING AND LEARNING OF STATISTICS
ASA NCTM
STatistics Education Web
STEW
online journal of K-12 statistics lesson plans

How long are the words in the Gettysburg Address?
Student Guide

The Gettysburg Address shown below was delivered by Abraham Lincoln on November 19, 1863 on the battlefield near Gettysburg, Pennsylvania. It is one of the most famous speeches ever given by an American President.

The Gettysburg 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 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.

In this investigation, we will address the question "How long are the words in the Gettysburg Address?"

1. a.   What are the population and the variable of interest for this investigation?

1. b.   Describe how would you select a sample of words to produce a sample that is similar to (representative of) the population?
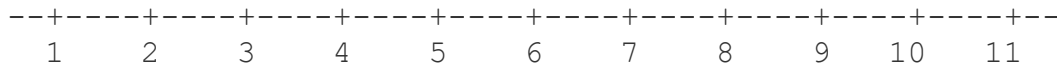
Part I – Self-Selected Sampling

The first method we will use to select a sample is called self-selection. With self-selection, use your eyes and circle 10 words you think are representative of the varying word lengths in the Gettysburg Address. Record the words and the lengths in Table 1.1.

Table 1.1. Self-Selected Sample of 10 Words (Sample 1)

Word Number	Word Selected	Length of Word
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
	Total	

2.   Use the plot below to make a dotplot of your sample data. Once completed, compare your dotplot with other students' dotplots. Are they the same or different?

```
--+----+----+----+----+----+----+----+----+----+----+--
  1    2    3    4    5    6    7    8    9   10   11
```

Dotplot 1.1. Length of Word (Sample 1)

Based on your dotplot, comment on the shape, center and variability of your sample data distribution.

3.   Did you use a strategy for selecting the words in your sample? Describe your strategy.

4. Determine the mean word-length for your sample (Add up the 10 word-lengths and divide by 10). Report your mean to the closest one-decimal place (for example, 7.1). Compare your mean with other students' means. Did you all get the same mean?

5. Using a large sheet of paper provided by your teacher, make a dotplot of the sample means for your class (Dotplot 1.2). Comment on the shape, center and variability of this distribution. Based on the dotplot, give a range of typical values for a sample mean using self-selected sampling.
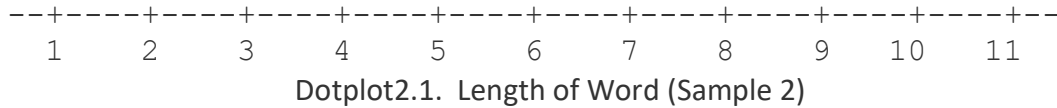
Part II – Simple Random Sampling

Another way to select a sample from a population is to use simple random sampling. How do we select a simple random sample of 10 words from the Gettysburg Address?

Your teacher will provide you with a list of the 268 words in the Gettysburg Address. Each word in this list has been assigned a number from 1 (001) to 268. To randomly select 10 words from the Gettysburg Address we will generate 10 distinct random integers between 1 and 268 and select the corresponding word from the Gettysburg Address. Your teacher will provide instructions on how to select 10 distinct random integers between 1 and 268. Once selected, enter your numbers in Table 2.1 below along with the corresponding words from the Gettysburg Address and the length of each word.

Table 2.1. Simple Random Sample of 10 Words (Sample 2)

Random Integer	Word Selected	Length of Word
	Total	

6.   Use the plot below to make a dotplot your sample data. Once completed, compare your dotplot with other students' dotplots. Are they the same or different?

```
--+----+----+----+----+----+----+----+----+----+----+--
  1    2    3    4    5    6    7    8    9   10   11
```
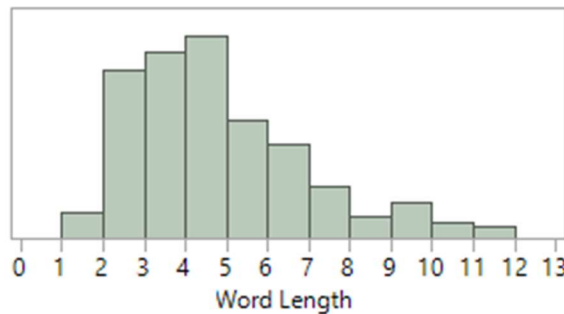Dotplot2.1. Length of Word (Sample 2)

Based on your dotplot, comment on the shape, center and variability of your sample data distribution.

7.   Determine the mean word-length for your sample (Add up the 10 word-lengths and divide by 10). Report your mean to the closest one-decimal place (for example, 7.1). Compare your mean with other students' means. Did you all get the same mean?

8.   Using a large sheet of paper provided by your teacher, make a dotplot of the sample means for your class (Dotplot 2.2). Comment on the shape, center and variability of this distribution. Based on the dotplot, give a range of typical values for a sample mean using simple random sampling.

9.   Compare Dotplots 1.2 (30 Means from Self-Selected Samples) and Dotplot 2.2 (30 Means from Simple Random Samples). Specifically, compare the centers and variability for these two distributions. Also, compare the ranges of typical values for a sample mean for each distribution.

10.   Which sampling method (self-selection or simple random) do you think is more likely to produce a sample representative of the population. Explain your reasoning.

11.   The histogram of word-lengths for the entire population of all 268 words in the Gettysburg Address is shown below.

Based on the histogram, comment on the shape, center and variability of the population distribution of word-lengths.

12.   The actual mean word length for all 268 words is approximately 4.3 characters. Let's compare your classes' sample means from each method with 4.3. Specifically, using self-selected sampling, how many sample means are below 4.3, equal to 4.3 and above 4.3? Using simple random sampling, how many sample means are below 4.3, equal to 4.3 and above 4.3? Now that you know this, which sampling method do you think would be better to use?

Summary of Parts I and II

Samples that are self-selected tend to produce *biased* results. In this investigation, in our self-selected samples, your eyes are drawn to the larger words and the means from the samples tend to overestimate the population mean. This is called *sampling bias*. With simple random sampling, randomness is introduced into the sampling procedure in order to produce samples that tend to be representative of the population.

Part III – Simple Random Sampling and Sample Size

In Part III, we will explore the relationship between sample size and how close a sample mean from a simple random sample is to the actual population mean.

Use a random number generator to select a simple random sample of 5 words from the Gettysburg Address. Enter your results into Table 3.1 along with the corresponding words from the Gettysburg Address and the length of each word.

Table 3.1. Simple Random Sample of 5 Words (Sample 3)

Word #	Word Selected	Length of Word
	Total	

13. Compute the mean for your sample. Using a large sheet of paper provided by your teacher, summarize how the means for your class vary in a dotplot (Dotplot 3.1). Compare the results displayed in Dotplot 2.2 (Means from samples of size 10) and Dotplot 3.1 (Means from samples of size 5). Specifically, compare the centers and variability for these two distributions. Also, compare the ranges of typical values for a sample mean for each distribution.

Part III of this lesson illustrates that with simple random samples, the larger the sample size, the closer sample means tend to be to the population mean.

Summary

When sampling words from the Gettysburg Address, self-selected samples tend over-represent longer words when compared to how often they occur in the population. Using simple random sampling tends to produce samples that are representative of the population. While the sample mean may not be the same as the population mean, we expect the sample mean to within a range of values centered around the population mean. Larger simple random samples tend to produce sample means closer to the population mean.

Gettysburg Address Word List (page 1)

Number	Word	Length		Number	Word	Length		Number	Word	Length
001	Four	4		046	Nation	6		091	Live.	4
002	Score	5		047	So	2		092	It	2
003	And	3		048	Conceived	9		093	Is	2
004	Seven	5		049	And	3		094	Altogether	10
005	Years	5		050	So	2		095	Fitting	7
006	Ago.	3		051	Dedicated,	9		096	And	3
007	Our	3		052	Can	3		097	Proper	6
008	Fathers	7		053	Long	4		098	That	4
009	Brought	7		054	Endure.	5		099	We	2
010	Forth	5		055	We	2		100	Should	6
011	Upon	4		056	Are	3		101	Do	2
012	This	4		057	Met	3		102	This.	4
013	Continent	9		058	On	2		103	But	3
014	A	1		059	A	1		104	In	2
015	New	3		060	Great	5		105	A	1
016	Nation:	6		061	Battlefield	11		106	Larger	6
017	Conceived	9		062	Of	2		107	Sense,	5
018	In	2		063	That	4		108	We	2
019	Liberty,	7		064	War.	3		109	Cannot	6
020	And	3		065	We	2		110	Dedicate,	8
021	Dedicated	9		066	Have	4		111	We	2
022	To	2		067	Come	4		112	Cannot	6
023	The	3		068	To	2		113	Consecrate,	10
024	Proposition	11		069	Dedicate	8		114	We	2
025	That	4		070	A	1		115	Cannot	6
026	All	3		071	Portion	7		116	Hallow	6
027	Men	3		072	Of	2		117	This	4
028	Are	3		073	That	4		118	Ground.	6
029	Created	7		074	Field	5		119	The	3
030	Equal.	5		075	As	2		120	Brave	5
031	Now	3		076	A	1		121	Men,	3
032	We	2		077	Final	5		122	Living	6
033	Are	3		078	Resting	7		123	And	3
034	Engaged	7		079	Place	5		124	Dead,	4
035	In	2		080	For	3		125	Who	3
036	A	1		081	Those	5		126	Struggled	9
037	Great	5		082	Who	3		127	Here	4
038	Civil	5		083	Here	4		128	Have	4
039	War,	3		084	Gave	4		129	Consecrated	11
040	Testing	7		085	Their	5		130	It,	2
041	Whether	7		086	Lives	5		131	Far	3
042	That	4		087	That	4		132	Above	5
043	Nation,	6		088	That	4		133	Our	3
044	Or	2		089	Nation	6		134	Poor	4
045	Any	3		090	Might	5		135	Power	5

Gettysburg Address Word List (page 2)

Number	Word	Length		Number	Word	Length		Number	Word	Length
136	To	2		181	Have	4		226	We	2
137	Add	3		182	Thus	4		227	Here	4
138	Or	2		183	Far	3		228	Highly	6
139	Detract.	7		184	So	2		229	Resolve	7
140	The	3		185	Nobly	5		230	That	4
141	World	5		186	Advanced.	8		231	These	5
142	Will	4		187	It	2		232	Dead	4
143	Little	6		188	Is	2		233	Shall	5
144	Note,	4		189	Rather	6		234	Not	3
145	Nor	3		190	For	3		235	Have	4
146	Long	4		191	Us	2		236	Died	4
147	Remember	8		192	Here	4		237	In	2
148	What	4		193	To	2		238	Vain,	4
149	We	2		194	Be	2		239	That	4
150	Say	3		195	Dedicated	9		240	This	4
151	Here,	4		196	To	2		241	Nation,	6
152	But	3		197	The	3		242	Under	5
153	It	2		198	Great	5		243	God,	3
154	Can	3		199	Task	4		244	Shall	5
155	Never	5		200	Remaining	9		245	Have	4
156	Forget	6		201	Before	6		246	A	1
157	What	4		202	Us,	2		247	New	3
158	They	4		203	That	4		248	Birth	5
159	Did	3		204	From	4		249	Of	2
160	Here.	4		205	These	5		250	Freedom,	7
161	It	2		206	Honored	7		251	And	3
162	Is	2		207	Dead	4		252	That	4
163	For	3		208	We	2		253	Government	10
164	Us	2		209	Take	4		254	Of	2
165	The	3		210	Increased	9		255	The	3
166	Living,	6		211	Devotion	8		256	People,	6
167	Rather,	6		212	To	2		257	By	2
168	To	2		213	That	4		258	The	3
169	Be	2		214	Cause	5		259	People,	6
170	Dedicated	9		215	To	2		260	For	3
171	Here	4		216	Which	5		261	The	3
172	To	2		217	They	4		262	People,	6
173	The	3		218	Gave	4		263	Shall	5
174	Unfinished	10		219	The	3		264	Not	3
175	Work	4		220	Last	4		265	Perish	6
176	Which	5		221	Full	4		266	From	4
177	They	4		222	Measure	7		267	The	3
178	Who	3		223	Of	2		268	Earth.	5
179	Fought	6		224	Devotion,	8				
180	Here	4		225	That	4				