

Stat 218 - Day 6
Random sampling

Very key terms:

Population- the entire collection of observational units in which one is interested

Sample- a (typically small) part of the population from which we collect information, in an effort to learn about the population

We want the sample to be **representative** of the population. How to achieve this?

Example: Elvis Lives?!

On the twelfth anniversary of the (alleged) death of Elvis Presley, a Dallas record company sponsored a national call-in survey. Listeners of over 1000 radio stations were asked to call a 1-900 number (at a charge of \$2.50) to voice an opinion concerning whether or not Elvis was really dead. It turned out that 56% of the callers felt that Elvis was alive.

(a) Identify the population and sample in this example.

(b) Do you think that 56% is an accurate reflection of beliefs of *all* Americans on this issue? If not, identify some of the flaws in the sampling method.

Example: President Landon?

In 1936, *Literary Digest* magazine conducted the most extensive (to that date) public opinion poll in history. They mailed out questionnaires to over 10 million people whose names and addresses they had obtained from phone books and vehicle registration lists. More than 2.4 million people responded, with 57% indicating that they would vote for Republican Alf Landon in the upcoming Presidential election. (Incumbent Democrat Franklin Roosevelt won the election, carrying 63% of the popular vote.)

(a) Identify the population and sample in this example.

(b) Suggest why the magazine's prediction was so much in error in Landon's favor.

Example: Famous Words

Consider the population of 268 words in the passage below.

- (a) Select a sample of ten representative words from this population.

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.

- (b) Record the number of letters in each word, and calculate the average number of letters per word.
- (c) Add your result to the graph on the board.
- (d) You will be told the average number of letters per word in the population of all 268 words. Do the sample averages from the class generally over- or under-estimate this value?

More Definitions

- **Biased sampling method**- one which systematically favors certain outcomes
- **Simple random sample**- every individual has the same chance of being selected, and every sample of its size has the same chance of being the sample actually selected

- (e) Use a table of random digits to take a simple random sample of five words from this population. (Use <http://www.rand.org/publications/classics/randomdigits/samples.txt>. Select three digits at a time, and disregard any numbers larger than 268. Also disregard repeats, and keep going until you have found five different three-digit numbers between 001 and 268.) Record the words chosen and the number of letters in each. Calculate the average number of letters per word in your sample, and add your result to the plot.

1 Four	35 in	69 dedicate	103 But,	137 add	171 here	205 these	239 that
2 score	36 a	70 a	104 in	138 or	172 to	206 honored	240 this
3 and	37 great	71 portion	105 a	139 detract.	173 the	207 dead	241 nation,
4 seven	38 civil	72 of	106 larger	140 The	174 unfinished	208 we	242 under
5 years	39 war,	73 that	107 sense,	141 world	175 work	209 take	243 God,
6 ago,	40 testing	74 field	108 we	142 will	176 which	210 increased	244 shall
7 our	41 whether	75 as	109 cannot	143 little	177 they	211 devotion	245 have
8 fathers	42 that	76 a	110 dedicate,	144 note,	178 who	212 to	246 a
9 brought	43 nation,	77 final	111 we	145 nor	179 fought	213 that	247 new
10 forth	44 or	78 resting	112 cannot	146 long	180 here	214 cause	248 birth
11 upon	45 any	79 place	113 consecrate,	147 remember,	181 have	215 for	249 of
12 this	46 nation	80 for	114 we	148 what	182 thus	216 which	250 freedom,
13 continent	47 so	81 those	115 cannot	149 we	183 far	217 they	251 and
14 a	48 conceived	82 who	116 hallow	150 say	184 so	218 gave	252 that
15 new	49 and	83 here	117 this	151 here,	185 nobly	219 the	253 government
16 nation:	50 so	84 gave	118 ground.	152 but	186 advanced.	220 last	254 of
17 conceived	51 dedicated,	85 their	119 The	153 it	187 it	221 full	255 the
18 in	52 can	86 lives	120 brave	154 can	188 is	222 measure	256 people,
19 liberty,	53 long	87 that	121 men,	155 never	189 rather	223 of	257 by
20 and	54 endure.	88 that	122 living	156 forget	190 for	224 devotion,	258 the
21 dedicated	55 We	89 nation	123 and	157 what	191 us	225 that	259 people,
22 to	56 are	90 might	124 dead,	158 they	192 to	226 we	260 for
23 the	57 met	91 live.	125 who	159 did	193 be	227 here	261 the
24 proposition	58 on	92 It	126 struggled	160 here.	194 here	228 highly	262 people,
25 that	59 a	93 is	127 here	161 It	195 dedicated	229 resolve	263 shall
26 all	60 great	94 altogether	128 have	162 is	196 to	230 that	264 not
27 men	61 battlefield	95 fitting	129 consecrated	163 for	197 the	231 these	265 perish
28 are	62 of	96 and	130 it,	164 us	198 great	232 dead	266 from
29 created	63 that	97 proper	131 far	165 the	199 task	233 shall	267 the
30 equal.	64 war.	98 that	132 above	166 living,	200 remaining	234 not	268 earth.
31 Now	65 We	99 we	133 our	167 rather,	201 before	235 have	
32 we	66 have	100 should	134 poor	168 to	202 us,	236 died	
33 are	67 come	101 do	135 power	169 be	203 that	237 in	
34 engaged	68 to	102 this.	136 to	170 dedicated	204 from	238 vain,	

- (f) Do the sample averages from the random samples generally over- or under-estimate the population average, or are they roughly split on both sides? Does the random sampling method appear to be unbiased?

To really examine the long-term patterns of random sampling, we will use technology to take many, many samples. From the course webpage, follow the links for JAVA applets, and select the “Sampling Words” applet. The information in the top right panels show the population distributions and tells you the average number of letters per word in the population, the population proportion of “long words,” and the population proportion of nouns. Unclick the boxes next to “Show Long” and “Show Noun” so we can continue to focus on the lengths of words for now.

- (g) Specify 5 as the sample size and click Draw samples. Record the lengths of the words and the average for the sample of 5 words.

Word	1	2	3	4	5	Avg
# letters						

- (h) Click Draw Samples again. Did you obtain the same sample of words this time? Did you obtain the same average length?
- (i) Change the Number of samples from 1 to 498. Click off the Animate button. Then click on the Draw Samples button. The applet now takes 498 more simple random samples from the population (for a total of 500) and adds the sample results to the graph in the lower right panel. The red arrow indicates the average of the 100 sample averages. Record this value below.
- (j) If the sampling method is unbiased the sample averages should be centered “around” the population average of 4.295 words. Does this appear to be the case?
- (k) What do you suspect would happen to the variability in sample averages if you took a random sample of 10 words rather than 5? Explain briefly.
- (l) Change the sample size in the applet to take 500 random samples of 10 words each. Summarize how the distributions of average word lengths compare in shape, center, and spread.
- (m) Would taking a larger sample with a biased sampling method help to reduce the bias? Explain.