Math 243

Day 16

Sampling Methods – Inv. 1.12

Review

So far, we've considered sampling from random processes:

Define π = probability of "success"

Calculate $\hat{p}$ from a sample of size n

Estimate π with a 95% CI or test $H_0: \pi = \#$

Requires that there is a "probability of success" that is the same for each trial.

Sampling from a Finite Population

Now suppose π is a population *proportion* and we wish to estimate or test π using data from a sample.

Q: How popular is K-TEC among people at OIT?

Define π to be the proportion of people at OIT who listen to K-TEC at least once a week

Population = people at OIT

Sample = people in Winter 2018 Math 243 on 02/07

$\hat{p}$ =

Plus Four 95%CI:

Generalizability

Do you believe that our class is representative of all people at OIT in their love of K-TEC?

If so, then believe the "Plus Four" CI for π.

If not, then you need a better method of obtaining a sample.

Sampling Methods

Choose a sample from the population by

- using the most readily available observational units: *convenience sample*

- giving each observational unit equal chance at being selected: *simple random sample (SRS)*

- dividing the population into homogenous groups then taking a SRS from each subgroup: *stratified random sample*
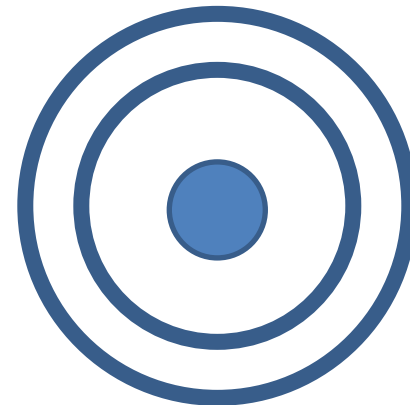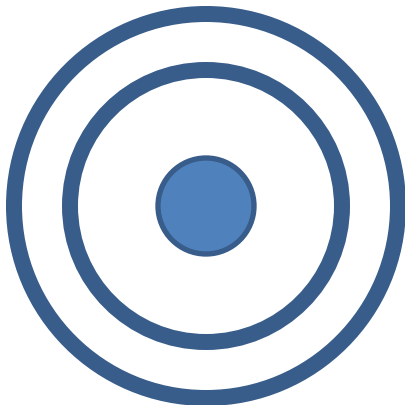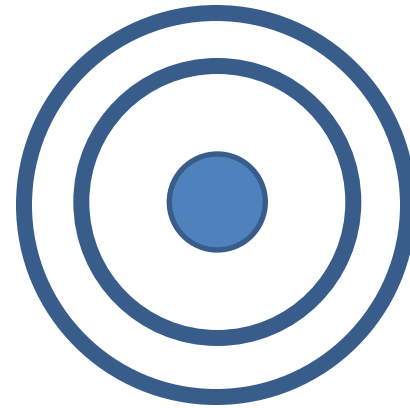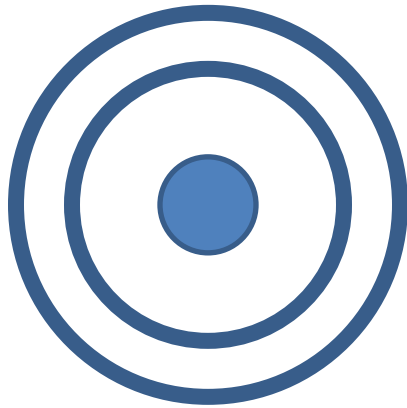
Examples of Sampling Methods

Suppose we want to choose a sample of 200 OIT students.

- *convenience sample =*

    10 Spring 2018 Math 243 students and the next 190 students I see on-campus who respond to my survey request

- *simple random sample (SRS) =*

    Obtain a list of all students enrolled at OIT from registrar's office, use random.org to randomly select 200 of them, then contact each of these 200 until they respond

- *stratified random sample =*

    Determine the proportion of OIT students in each department, then take a SRS from each department of proportional size (e.g. if 1.5% of OIT students are math majors, then we would randomly select 0.015*200 = 3 math majors for our sample, etc.)

What is a "good" sampling method?

Think about the population parameter as the target and the sample as your shots...

What is a "good" sampling method?

Unbiased – the samples don't systematically differ from the population in the characteristic of interest, that is, **on average,** the sample statistics are close to the population parameter

Precise – the sample statistics tend to be close to each other.

Inv. 1.12

Population: 268 words of Gettysburg Address

Suppose we want to estimate the **average** length of a word in the address, **μ**

Part (a): Pick a representative sample of 10 words and compute the sample average, $\bar{x}$.

(a) Circle 10 representative words in the following passage.

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.

(g) Is your sample average similar to the population average? Did everyone in your class obtain the same sample average? Describe a way for deciding whether this sampling method tends to produce samples that are representative of the larger population.

(h) Construct a dotplot displaying the *average length of words* in your sample and those of your classmates.

```
├────┬────┬────┬────┬────┬────┬────┬────┬────┬────┬────┬────
0    1    2    3    4    5    6    7    8    9    10   11
```
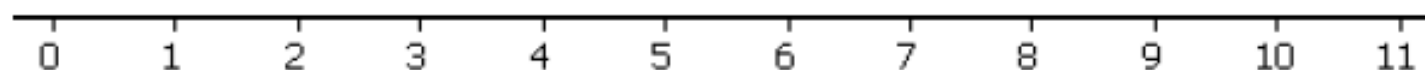
Label:

What are the observational units in this graph?

Inv. 1.12

Population: 268 words of Gettysburg Address

Suppose we want to estimate the average length of a word in the address, μ

Method 1. Pick a sample of 10 words and compute the sample average, $\bar{x}$. (*convenience sample*)

Method 2. Use the random applet to pick a sample of 5 words (*SRS*)

- Are student selections of 10 words unbiased? Part (h)
- Are SRS of 5 words unbiased? Part (o)

(m) Select a simple random sample of 5 ID numbers from the sampling frame. Match these ID numbers with the corresponding words. Write down the resulting ID numbers, the selected words, and the length of each word.

	1	2	3	4	5
ID number					
Word					
Length					

(n) Calculate the average word length in your random sample of 5 words.

(o) Again combine your results with your classmates to produce a dotplot of the sample mean word lengths. Comment on the distribution and particularly how it compares to the one in (h).

```
    0    1    2    3    4    5    6    7    8    9    10   11
              Label:
```
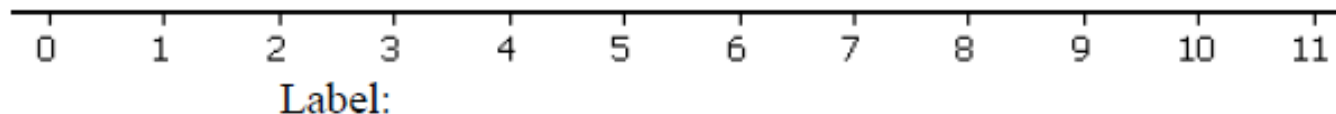
Comparison:

Sampling Frame – a *list* of all objects or people in the population

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,	

Generate Random Numbers from Applet

Generate Random Numbers

Number of replications:	1	30,155,160,185,235
Numbers per replication:	5	
Number range:	From: 1	
	To: 268	
With replacement?	No ▾	
Sort the results?	Yes: 1,2,3 ▾	
	Generate	

Sampling methods and generalizability

If you're sampling from a finite population in order to estimate a **population proportion**,

then you need to use a **random** sampling method (SRS, stratified)

If you're sampling from a Binomial process in order to estimate a **probability of success**,

then you must sample **independent trials under identical conditions**

Convenience samples are rarely, if ever, useful for generalizing to a larger group.