Unit 9: Sampling

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 臣 の�?

Odd ex: Know population, guess sample!

The 10 million people in a state own an avg of 10 shirts (or blouses, or tops, or ...), with an SD of 4, but 3 million own at least 12. In a simple random sample of 1600:

(a) What is the probability that the avg # of shirts of people in sample is at least 12?

(a) Box values are numbers of shirts.

- EV of avg = 10,
- SE of avg = $4/\sqrt{1600} = .1$,
- so 12 in std units is z = (12 10)/.1 = 20
- $P(z \ge 20)$ is negligible.

Odd ex: Know population, guess sample!

The 10 million people in a state own an avg of 10 shirts (or blouses, or tops, or \ldots), with an SD of 4, but 3 million own at least 12. In a simple random sample of 1600:

(b) What is the probability that at least 500 people in sample own at least 12 shirts?

(b) Box is 30% 1's, 70% 0's.

- EV of count = 1600(.3) = 480,
- SE = $\sqrt{(.3)(.7)}\sqrt{1600} \approx 18.3.$
- ► $P(\text{count} \ge 500) = P(z \ge (500 480)/18.3 \approx 1.09) \approx ((100 72)/2)\% = 14\%.$

When the fraction p of a 0-1 distribution is close to 0% or 100%, the standard deviation σ , given by $\sqrt{p(1-p)}$, is small; but this effect becomes noticeable only very close to 0% or 100%. In particular, σ is at least half its maximum value of .5 except for p < 7% or p > 93%.



_								~ •		
	D7 = =IF(AND(B\$7>C7-0.2*SQRT(C7*(1-C7)),B\$7 <c7+0.2*sqrt(c7*(1-c7))),1,0)< th=""></c7+0.2*sqrt(c7*(1-c7))),1,0)<>									
	A	В	С	D	E	F	G	Н		
1	lf a sample is	nearly all 05	s or all 1's, a 9	5% confidence	interval can't					
2	be found: the	usual interva	al doesn't inclu	de the true % o	f 1's often					
3	enough (i.e., 95% of the time).									
4										
5	Pop: 1099	Real % 1's	Avgs of smpls	ls real %	What % of c.i.'s					
6	(~ 2% 1's)	in pop	(100 each)	in 95% c.i.?	include real %?					
7	0	0.0117188	0	0	72.5%					
8	0		0	0						
9	0		0	0						
10	0		0	0						
11	0		0	0						
12	0		0	0						
13	0		0.01	1						
14	0		0.01	1						
15	0		0.01	1						
16	0		0.01	1						
17	0		0.01	1						
18	0		0.01	1						
19	0		0.01	1						
20	0		0.01	1						
21	0		0.01	1						
22	0		0.01	1						
23	0		0.01	1						
24	0		0.01	1						
25	0		0.01	1						

							_	
D7 = = = IF(AND(B\$7>C7-0.2*SQRT(C7*(1-C7)),B\$7 <c7+0.2*sqrt(c7*(1-c7))),1,0)< th=""><th>1,0)</th></c7+0.2*sqrt(c7*(1-c7))),1,0)<>								1,0)
	A	В	C	D	E	F	G	Н
1	lf a sample is	closer to ha	lf O's and half 1	's, the formula	for a 95%			
2	confidence int	terval is muc	h more reliable	, i.e., includes	the true %			
3	of 1's more fre	equently (tho	ugh not always	s 95% of the tin	ne).			
4								
5	Pop: 1099	Real % 1's	Avgs of smpls	ls real %	What % of c.i.'s			
6	(~ 40% 1's)	in pop	(100 each)	in 95% c.i.?	include real %?			
7	0	0.4101563	0.49	1	94.4%			
8	1		0.49	1				
9	0		0.49	1				
10	1		0.49	1				
11	0		0.49	1				
12	0		0.49	1				
13	1		0.5	1				
14	0		0.5	1				
15	1		0.51	1				
16	0		0.5	1				
17	1		0.51	1				
18	0		0.5	1				
19	0		0.5	1				
20	0		0.51	1				
21	1		0.51	1				
22	1		0.51	1				
23	0		0.5	1				
24	1		0.5	1				
25	1		0.49	1				

How large a sample needed? (I)

For %, estimate p and solve for n:

• Because tolerance $= z \sqrt{p(1-p)/n}$ (get z from normal table), ...

•
$$n = z^2 p(1-p)/\text{tolerance}^2$$

Ex: How many people should we survey to get % of voters in favor of Smith within 2% (95% CI)? Probably about 20% favor Smith.

... or just note
$$p(1-p) \leq 0.25$$
:

How large a sample needed? (II)

For estimates of μ , <u>must</u> guess *s* (because we can't get *s* from μ) Again, solve for $n : z^2\sigma^2$ /tolerance²

Ex: To estimate within 1 pt (95% CI) the population's avg score on an exam, how big a sample is needed? Guess $\sigma = 10$.

(Answer: About 400.)