

LECTURE 15: CONFIDENCE INTERVALS III

I. CIs for Proportions

- a. A proportion measures the fraction of a population, such as the percent of female viewers, the portion of customers who use a coupon, or the fraction of voters who will vote for Gandalf the Grey.
 - i. Proportions cannot be less than 0 or greater than 1.
- b. The equation for a CI should look familiar:

$$\widehat{CI}_{\bar{p}} = \bar{p} \mp z_{\alpha/2} \hat{\sigma}_p = \bar{p} \mp z_{\alpha/2} \sqrt{\frac{\bar{p}(1 - \bar{p})}{n}}$$

- i. Where *p-bar* is the sample proportion of the population; and
- ii. *n* is the sample size.
- iii. Note we use the z-score here.

II. Example

- a. A polling company wants to know what percent of voters will support a new ballot initiative in the upcoming election. They interview 500 voters and find 56% support the initiative. At 99% confidence, what is the confidence interval of the poll?
- b. Sadly, Excel does not have this equation built in (yet) so we have to do this the old fashion way. Set up Excel like this:

	A	B
1	p-bar	0.56
2	n	500
3	z	2.576
4		
5	MoE	
6		
7	LB	
8	UB	

- i. Where MoE is the margin of error, LB is the lower bound, and UB is the upper bound.
- ii. In B5, we do this: =B3*SQRT(B1*(1-B1)/B2)
- iii. Keep in mind that SQRT is “square root.” We are taking the square root of 0.56 times 0.44 divided by 500. Then we multiply the whole thing by the z score. You should get about 0.0572.

- iv. For the lower and upper bounds, add and subtract from p-bar, as you've done previously with x-bar:

	A	B
1	p-bar	0.56
2	n	500
3	z	2.576
4		
5	MoE	=B3*SQRT(B1*(1-B1)/B2)
6		
7	LB	=B1-B5
8	UB	=B1+B5

- v. So you should get:

	A	B
1	p-bar	0.56
2	n	500
3	z	2.576
4		
5	MoE	0.0572
6		
7	LB	0.5028
8	UB	0.6172

- c. There's a 99% chance that the true support of voters is between about 50.3% and 61.7%.

III. Additional Notes on Confidence Intervals

- a. Choosing a confidence level is tricky, which is why sometimes you see multiple levels reported. That's not always the case, though, because—bizarrely—people often want definite answers when it comes to statistics.
- On one hand, a narrow range tells you a lot about what the population mean might be. You are more precise but risk completely missing the mark.
 - On the other hand, a wider range ensures the population mean is in that band. Your range probably includes the parameter, but you're vague.
 - The question is, what side is best to err on? That changes with circumstance. That said, 95% confidence is quite low: never go lower and usually go higher.
 - Sometimes you see people refer to a 90% level. I used to give those z values but took them out because 90% is frankly pretty desperate.