

Chapter 19: CIs for Proportions

1. In a recent Gallup Poll (<http://www.gallup.com/poll/104464/Economy-Surpasses-Iraq-Most-Important-Problem.aspx>), 34% of Americans mention the economy as the most important problem facing the country. Results were based on a telephone interview of 1007 national adults.
 - a) Check if the assumptions/conditions are met.
 1. Independence –
 - a. Randomization – It is reasonable to assume that the adults were randomly selected.
 - b. 10% Condition – 1007 is less than 10% of all national adults
 2. Success/Failure – $np = 1007 * .34 = 342 \geq 10$. Also, $nq \geq 10$.
 - b) Construct a 95% confidence interval.
(0.310373, 0.368873)
 - c) Interpret the confidence interval in the context of this problem.
We are 95% confident that between 31% and 37% of national adults feel that the economy is the most important problem facing the country.
 - d) What does “95% confidence” mean? Pick one for each set of parentheses.
If we select repeated random samples of the (**SAME SIZE**, XXXXXXXXXXXXXXX, XXXXXXXXXXXXXXX), we would expect 95% of the confidence intervals to contain the (**TRUE**, XXXXXX) proportion of national adults who mention the economy as the most important problem facing the country.
 - e) A politician claims that “the majority of Americans think the economy is the most important problem facing the country”. Does the interval support or contradict the politician? Explain.
Contradicts. The intervals shows that at most 37% feel that way, not the majority.
2. In Fall 2009, I claimed to collect homework 25% of the time. Between the two classes, I had collected homework 8 times (3 in one and 5 in the other). There were 30 times that I might have collected homework.
 - a) Indicate why the assumptions/conditions for constructing a confidence interval are NOT met.
The success/failure condition is not met. We only have 8 successes.
 - b) In a simulation with 50 trials, there were 13 “hw” and 37 “noHW”.
 - i) Show that the assumptions/conditions are now met.
 1. Independence – The trials were independent by construction. This also allows us to ‘pass’ the randomization and 10% conditions.
 2. Success/Failure – $np = 13 \geq 10$. Also, $nq = 37 \geq 10$.
 - ii) Construct a 95% confidence interval.
(0.138419, 0.381581)
 - iii) Interpret the confidence interval in the context of this problem.
We are 95% confident that homework is collected between 14% and 38%.
 - iv) Based on the interval you constructed, is 25% plausible for the amount of time I collect HW?
Yes. The interval contains 25%.
 - v) The interval obtained is quite large. What can we do to make the interval smaller? (*Hints: Think sample size and level of confidence...*)
 - Increase sample size (which decreases the standard error, which decreases the margin of error, which decreases the interval width)
 - Decrease the level of confidence (which decreases the interval width)
 - c) Now you will make your own simulation by creating your own sample in Minitab. Since we need numerical values, enter values 0 and 1 to indicate “noHW” and “HW” in a column. Enter the correct probabilities in another column. Use Random Data > Discrete Distribution to create a sample of 1000 observations.
 - i) Why can we not use “Sample From Columns”?
That would sample as if “noHW” and “HW” were equally likely, which they aren’t.
 - ii) Construct a 95% confidence interval (in the 1P dialog box use “Samples in columns” instead of “summarized data”) and comment on the difference between these answer and the interval in (ii) above.

My Data and Interval:

Variable	X	N	Sample p	95% CI
C3	228	1000	0.228000	(0.201997, 0.254003)

Based on my interval, we are 95% confident that homework will be collected between 20.2% and 25.4% of the time. Note that my claimed 25% is still plausible.

Many of you will also come to this conclusion, but some won't. This is to be expected and is precisely what is meant by the "95% confident". In fact, approximately 95% of you will draw the same conclusion I did, but 5% of you won't.