

USING STATCRUNCH TO CONSTRUCT CONFIDENCE INTERVALS and CALCULATE SAMPLE SIZE

Using StatCrunch for confidence intervals (CI's) is super easy. As you can see in the assignments, I cover 9.2 before 9.1 because I feel processes involving means are more intuitive. If at any time with any of these processes you're given data instead of summary stats, use the data. It's so much easier. You just choose "With data" instead of "With summary". StatCrunch will calculate all the summary statistics the process needs and automatically plug them in.

This section, 9.2 requires us to use a new distribution, T. T makes up the extra error that we accumulate by not knowing the value of the population standard deviation (How would we know it if we don't know the mean?) and instead, estimating it with the sample standard deviation. Here are a couple of problems from 9.2.

31. How Much Do You Read? A recent Gallup poll asked 1006 Americans, "During the past year, about how many books, either hardcover or paperback, did you read either all or part of the way through?" Results of the survey indicated that $\bar{x} = 13.4$ books and $s = 16.6$ books. Construct a 99% confidence interval for the mean number of books that Americans read either all or part of during the preceding year. Interpret the interval.

32. How Much Do You Read? A Gallup poll conducted July 21–August 14, 1978, asked 1006 Americans, "During the past year, about how many books, either hardcover or paperback, did you read either all or part of the way through?" Results of the survey indicated that $\bar{x} = 18.8$ books and $s = 19.8$ books.

- (a) Construct a 99% confidence interval for the mean number of books Americans read either all or part of during the preceding year. Interpret the interval.
- (b) Compare these results to those of Problem 31. Were Americans reading more in 1978?

In #31, we want to construct and interpret a 99% CI. We have the sample size, $n = 1006$, the sample standard deviation, $s = 16.6$ and the sample mean, $\bar{x} = 13.4$. That's everything we need. Open StatCrunch and click "Stat", "T stats", "One sample", "With summary" because we're given the summary statistics. Type in the value of the sample mean, etc. Click the button in front of "Confidence interval for μ " and enter your confidence level. Here's what you should have.

One Sample T Summary

Sample mean: 13.4
Sample std. dev.: 16.6
Sample size: 1006

Perform:

Hypothesis test for μ
Ho: $\mu = 0$
HA: $\mu \neq 0$

Confidence interval for μ
Level: 0.99

Output:
 Store in data table

? Cancel Compute!

Click "Compute" and you should get this output.

Options					
99% confidence interval results:					
μ : Mean of population					
Mean	Sample Mean	Std. Err.	DF	L. Limit	U. Limit
μ	13.4	0.52337033	1005	12.049322	14.750678

The lower and upper bounds of the CI are always on the right side of the output under “L. Limit” and “U. Limit”. Rounding to one decimal place, we have 12.0 to 14.8. To interpret, we say that “We are 99% confident that the *population mean* (μ !) number of books Americans read either all or part of the preceding year is between 12.0 and 14.8.” Note that I got this interpretation straight from the problem. Number 32 is the same thing, but the data is from 1978. Note that you’re being asked to find A CI FOR A MEAN. Therefore, T interval. You try, then compare with my answer.

Options					
99% confidence interval results:					
μ : Mean of population					
Mean	Sample Mean	Std. Err.	DF	L. Limit	U. Limit
μ	18.8	0.62426099	1005	17.188951	20.411049

This means that “We are (or I am) 99% confident that the *population mean* number of books Americans read either all or part of in 1978 is between 17.2 and 20.4.”

Part b wants us to compare the two intervals and asks if Americans were reading more books in 1978. If you either graph both intervals on a number line or position them relative to one another, note that there is no overlap of the two and the entire interval from 1978 is larger than the entire interval from last year. Therefore, we can conclude that it appears that Americans were reading more books in 1978.

Also from 9.2, finding sample sizes. StatCrunch has added a feature that makes this a lot easier than it used to be. If you’re following my class notes, that process will work but this is easier.

Before a study is done, we would generally set a goal in terms of confidence level and margin of error we’d be willing to accept, and calculate what sample size will allow us to meet that goal.

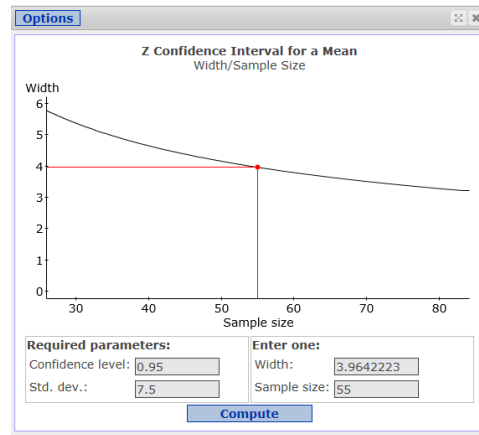
48. Television A researcher wanted to determine the mean number of hours per week (Sunday through Saturday) the typical person watches television. Results from the Sullivan Statistics Survey indicate that $s = 7.5$ hours.

- How many people are needed to estimate the number of hours people watch television per week within 2 hours with 95% confidence?
- How many people are needed to estimate the number of hours people watch television per week within 1 hour with 95% confidence?
- What effect does doubling the required accuracy have on the sample size?
- How many people are needed to estimate the number of hours people watch television per week within 2 hours with 90% confidence? Compare this result to part (a). How does decreasing the level of confidence in the estimate affect sample size? Why is this reasonable?

First of all, note how the problem is worded. “HOW MANY PEOPLE ARE NEEDED...” so they want sample size. The “within 2 hours” here is the margin of error (ME or just E) which is half the length of the CI. It indicates that we want our entire interval to be within 4 total units wide. The ME in the problem will always be given as “within x units...”. Ok, so we have

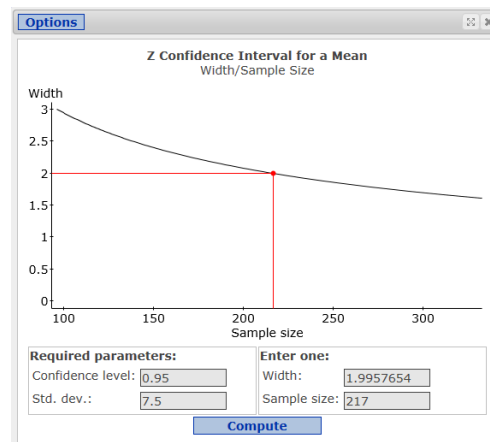
$s = 7.5$, $ME = 2$ and we want 95% confidence. Note that this is a Z procedure because each T distribution is defined by its degrees of freedom, which is the sample size minus one and sample size is what we're looking for! In StatCrunch, click "Stat", "Z stats", "One sample", "Width/sample size", type in all the relevant stuff and "Compute". It's really important here to remember that StatCrunch wants the ENTIRE LENGTH of the interval, not just the ME so you must DOUBLE the ME.

Here's what the output looks like.

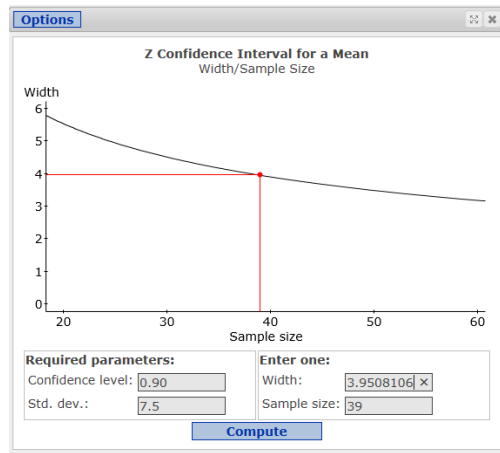


Note that you can drag the red dot on the graph and see how the width of the CI affects the sample size. Kind of fun! See how the width you entered (4) is now 3.964...? It means that at $n = 55$, you will get an interval under 4 total units wide. Go back to StatCrunch and change the sample size from 55 to 54. It'll show that you will be over your 4 unit goal if you reduce the sample size.

In part b, check out what happens when you halve the ME, thus halving the entire interval length.



It approximately quadruples the sample size needed! In part d, what happens if we leave our ME at 2 but decrease the confidence level to 90%?



We would only need to sample 39 people. Less confidence, smaller sample size needed. There's only one thing a person can do to both have a high level of confidence AND a small-ish margin of error and that's increase the sample size.

Section 9.1. Note that you don't see anything about means, averages, standard deviations in these problems. What you do see is either a percent or n people/items sampled with x people/items in the sample meeting a criteria.

25. Smartphone Purchases In a survey of 3611 adult Americans 18 years and older conducted in July 2010 by SmartRevenue, it was found that 542 have used their smartphone to make a purchase.

- (a) Obtain a point estimate for the population proportion of adult Americans 18 years and older who have used their smartphone to make a purchase.
- (b) Verify that the requirements for constructing a confidence interval about p are satisfied.
- (c) Construct a 90% confidence interval for the population proportion of adult Americans who have used their smartphone to make a purchase.
- (d) Interpret the interval.

Here, we have 3611 adult Americans sampled and 542 have used their smartphone to make a purchase (the criteria). Therefore, $x = 542$ and $n = 3611$.

In part a, we want a point estimate. In this case, the point estimate is the value of \hat{p} which is $x/n = 542/3611$, about 0.15.

I'm not going to do part b, but all it wants is for us to verify that the assumptions, or criteria for the process are met. That would be that $np(1-p)$ is at least 10 and that the sample size, n is less than 5% of the population size, N . These two assumptions ensure that our sample is random and independent.

Part c is asking for a confidence interval. This is ridiculously easy. "Stat", "Proportion stats", "One sample", "With summary". All you have to do is type in the x and n , click in front of "Confidence interval for p ", be sure to include the correct confidence level and "Compute" like this.

Your output should look like this.

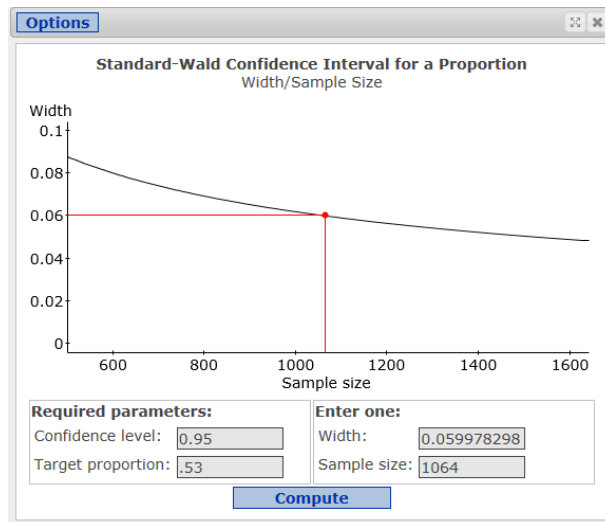
Proportion	Count	Total	Sample Prop.	Std. Err.	L. Limit	U. Limit
p	542	3611	0.15009693	0.0059436996	0.14032041	0.15987344

Again, the limits of the CI are on the right in the output and in this case, 0.14 to 0.16. You should get decimals, since proportions/percentages are numbers between 0 and 1. Interpret: “We are 90% confident that the *population proportion* of adult Americans 18 and older who have used their smartphone to make a purchase is between 0.14 and 0.16”. Or you can say “...between 14% and 16%”. Same thing.

Again in 9.1, we want to find the sample size needed to estimate a proportion. Note the problem reads “... estimate the proportion...” and “What sample size should be obtained...”. Thus, sample size to estimate a proportion. There are two possibilities. From the CI formula, if you set the margin of error part equal to E and solve it for n, the formula includes \hat{p} which we’d need the sample size, n to calculate! And n is the thing we’re looking to find! There are two ways of dealing with this. First, there are many times when estimations are updated to reflect the current population. When this happens, it’s okay to use the \hat{p} from that prior study. In part a below, there was a study in 2010 that found \hat{p} to be 0.53 which we can use. In StatCrunch, it’s called the “Target proportion”. Note that the ME is “3 percentage points” or 3% or 0.03. Again, you’ll need to double that for StatCrunch to a width of 0.06.

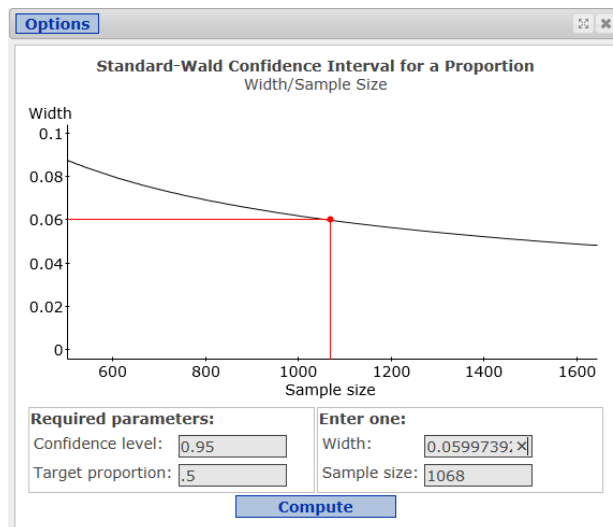
- 37. Football Fans** A television sports commentator wants to estimate the proportion of Americans who “follow professional football.” What sample size should be obtained if he wants to be within 3 percentage points with 95% confidence if
- (a) he uses a 2010 estimate of 53% obtained from a Harris poll.
 - (b) he does not use any prior estimates?
 - (c) Why are the results from parts (a) and (b) so close?

Go to “Stat”, “Proportion stats”, “One sample”, “Width/Sample size”. After typing in the confidence level of 95%, the width and \hat{p} estimate, you should have this.



And the sample size needed to meet the criteria in the problem is 1064 Americans.

Part b is similar, except we assume we have no prior estimate of \hat{p} . When this happens, we have to assume the worst case scenario and max out the value of \hat{p} , which happens to be 0.5. Therefore, the target proportion when we have no prior estimate of \hat{p} is 0.5. You use the same tool in StatCrunch that you did for part a. Here's what you should get. Note that the sample size, at 1068 is only four more than if you have a prior estimate. That's because the prior estimate, 0.53 is so close to the max of 0.5 which is the answer to part c.



The last example I have for you is from 9.3, estimating a population standard deviation. This is a little tricky. You need to remember that the variance and the standard deviation are both measures of variability and that the variance is the square of the standard deviation. The standard deviation then, is the principle square root of the variance. The tool for this process in StatCrunch is inference for a VARIANCE, not standard deviation (which is common) so if the question asks you for a CI for a variance, what you get from StatCrunch is exactly what you need. If you're asked to find a CI for a standard deviation, you'll need to take the square root of the lower and upper bounds. Here's an example.

11. 4 GB Flash Memory Card The following data represent the prices for a 4 GB flash memory card from online retailers. Treat these data as a simple random sample of all online retailers. In Problem 35 from Section 9.2, it was verified that the data are normally distributed and that $s = 2.343$ dollars. Construct and interpret a 95% confidence interval for the standard deviation price of a 4 GB flash memory card. Interpret the interval.

12.25	13.49	13.76	14.75	14.99
14.99	15.27	15.95	17.88	20.49

Source: ScanLife

I think all the problems in this section of the textbook give data, not summary stats. But we'll work off of the summary stats here. As usual, if you're given data in your problem, it's much easier to just use the data.

Note that the sample size is $n = 10$, the sample standard deviation is $s = 2.343$ and the confidence level is 95%. Because we'll use inference for a variance, we first square the standard deviation, so the sample variance is $s^2 = 5.489649$. I'm using all of the decimal places because you don't want to keep rounding and using rounded numbers over and over. Your answer will be off. Go to "Stat", "Variance stats", "One sample", "With summary". Type in your sample statistics, click the button in front of "Confidence interval for σ^2 " and be sure to include the correct confidence level. You should have this.

When you "Compute", you get this. Note that it says right on the output, " σ^2 : Variance of population". That means that "We are 95% confident that the *population variance* price of a 4 GB flash memory card is between \$2.60 and \$18.30". But that's not what we're asked to find.

95% confidence interval results:				
σ^2 : Variance of population				
Variance	Sample Var.	DF	L. Limit	U. Limit
σ^2	5.489649	9	2.5972478	18.296191

To get the CI for a population standard deviation, we take the square root of both the upper and lower bounds. Again, be careful not to round too much yet. $\sqrt{2.59724} = 1.61159\dots$ and $\sqrt{18.29619} = 4.27740\dots$ NOW we round and can say

that “We are 95% confident that the *population standard deviation* price of a 4 GB flash memory card is between \$1.61 and \$4.28”.