Calculating Confidence Intervals around Standard Deviations.

If we calculate a standard deviation (SD) from a sample, this value is an estimate of the true value in the population. In small samples, our estimate can be quite far off. But due to the law of large numbers, as our sample size increases, we will be measuring the standard deviation more accurately. Since the sample standard deviation is an estimate with uncertainty, we can calculate a 95% confidence interval around it.

Expressing the uncertainty in our estimate of the standard deviation can be useful. When researchers plan to simulate data, or perform an a-priori power analysis, they need accurate estimates of the standard deviation. For simulations, the standard deviation needs to be accurate because we want to generate data that will look like the real data we will eventually collect. For power analyses, we often want to think about the smallest effect size of interest (SESOI), which can be specified as the difference in means which you care about. To perform a power analysis, we also need to specify the expected standard deviation of the data. Sometimes researchers will use pilot data to get an estimate of the standard deviation. Since the estimate of the population standard deviation based on a pilot study has some uncertainty, confidence intervals might be a useful way to quantify the amount of uncertainty.

Open the R code Confidence_Intervals_for_Standard_Deviations.R to calculate the confidence interval around a standard deviation from a sample. You can also use this free GraphPad calculator to calculate confidence intervals for standard deviations. Run lines 1 to 3 to load the pwr package. Lines 5 to 8 specify the alpha level (and thus the 1-α confidence interval), the number of observations (n), the population standard deviations (st_dev), and the smallest effect size a researcher is interested in (SESOI). The code calculates the critical values for the upper and lower confidence interval (lines 11 and 12), and then calculates the confidence interval for the standard deviation (lines 15 and 16).

Q1: If you run lines 5 to 16 you will see that with an alpha level of 0.05, 100 observations, and a true standard deviation of 1, the 95% CI is [0.88; 1.16]. **Change the assumed population standard deviation from 1 to 2** in line 7. Keep all other settings the same. What is the 95% CI around the standard deviation of 2 with 100 observations (run lines 5 to 16)?

A) 95% CI [1.38; 3.65]

B) 95% CI [1.76; 2.32]

C) 95% CI [1.82; 2.22]

D) 95% CI [1.84; 2.20]

Q2: **Change the assumed population standard deviation back from 2 to 1** in line 7. **Lower the sample size from 100 to 20** in line 6. This will inform us about the width of the confidence interval for a standard deviation when we run a pilot study with 20 observations. Keep all other settings the same. What is the 95% CI around the standard deviation of 1 with 20 observations (run lines 5 to 16)?

A) 95% CI [0.91; 1.11]

B) 95% CI [0.82; 1.28]

C) 95% CI [0.76; 1.46]

D) 95% CI [1.52; 2.92]

Q3: If we want the 95% CI around the standard deviation of 1 to be at most 0.05 away from the assumed population standard deviation, how large should our number of observations be? Note that this means we want the 95% CI to fall within 0.95 and 1.05. But notice from the calculations above that the distribution of the sample standard deviations is not symmetrical. Standard deviations can't be smaller than 0 (because they are the square rooted variance). So in practice the question is: What is the smallest number of observations for the upper 95% CI to be smaller than 1.05? Replace n on line 6 with each of these values.

A) n = 489

B) n = 498

C) n = 849

D) n = 948

Let's explore what the consequences of an inaccurate estimate of the population standard deviation are on a-priori power analyses. Let's imagine we want to perform an a-priori power analysis for a smallest effect size of interest of half a scale point (on a scale from 1-5) on a measure that has an (unknown) true population standard deviation of 1.2.

Q4: **Change the number of observations in line 6 to 50**. **Change the assumed population standard deviation in line 7 to 1.2**. Keep the SESOI in line 8 to 0.5. The 95% confidence interval for the standard deviation based on a sample of 50 observation ranges from 1.002 to 1.495 (run lines 5 to 16 to confirm). To perform an a-priori power analysis we need to calculate Cohen's d, which is the difference divided by the standard deviation. In our example, we want to at least observe a difference of 0.5. What is Cohen's d (SESOI/SD) for the lower bound of the 95% confidence interval (where SD = 1.002) or the upper bound (where SD = 1.495)? **You can perform these calculations by running lines 19 and 21**.

A) d = 0.33 and d = 0.50

B) d = 0.40 and d = 0.60

C) d = 0.43 and d = 0.57

D) d = 0.29 and d = 0.55

If we draw a sample of 50 observations we can happen to observe a value that, due to random variation, is much smaller or much larger than the true population value. We can examine the effect this has on the number of observations that we think will be required when we perform an a-priori power analysis.

Q5: An a-priori power analysis is performed that uses the estimate of Cohen's d based on the lower 95% CI of the standard deviation. Which statement is true?

A) Because the lower bound of the 95% CI is **smaller** than the true population SD, Cohen's d is **smaller**, and the a-priori power analysis will yield a sample size that is **smaller** than the sample size we really need.

B) Because the lower bound of the 95% CI is **smaller** than the true population SD, Cohen's d is **larger**, and the a-priori power analysis will yield a sample size that is **larger** than the sample size we really need.

C) Because the lower bound of the 95% CI is **smaller** than the true population SD, Cohen's d is **smaller**, and the a-priori power analysis will yield a sample size that is **larger** than the sample size we really need.

D) Because the lower bound of the 95% CI is **smaller** than the true population SD, Cohen's d is **larger**, and the a-priori power analysis will yield a sample size that is **smaller** than the sample size we really need.

Q6: Let's check if our answer on the previous question was correct. We still have an alpha level of 0.05, n = 50, a standard deviation of 1.2, and a SESOI of 0.5. Run lines 23 and 24. The first power analysis uses Cohen's d based on the lower bound of the 95% confidence interval. The second power analysis uses the upper bound of the 95% confidence interval. (There is also a third power analysis based on the (in real-life situations unknown) true standard deviation, just for comparison, on line 27). Which statement is true?

A) The sample size per group is 68 when calculating the effect size based on the lower bound on the 95% CI around the standard deviation, and 86 when using the upper bound of the 95% CI around the standard deviation.

B) The sample size per group is 68 when calculating the effect size based on the lower bound on the 95% CI around the standard deviation, and 123 when using the upper bound of the 95% CI around the standard deviation.

C) The sample size per group is 86 when calculating the effect size based on the lower bound on the 95% CI around the standard deviation, and 123 when using the upper bound of the 95% CI around the standard deviation.

D) The sample size per group is 86 when calculating the effect size based on the lower bound on the 95% CI around the standard deviation, and 189 when using the upper bound of the 95% CI around the standard deviation.

It is clear sample sizes from a-priori power analyses depend strongly on an accurate estimate of the standard deviation. Keep into account that **estimates of the standard deviation have uncertainty**. Use **validated or existing measures for which accurate estimates of the standard deviation in your population of interest are available**, so that you can rely on a better estimate of the standard deviation in power analyses.

Some people argue that **if you have such a limited understanding of the measures you are using that you do not even know the standard deviation of the measure in your population of interest, you are not ready to use that measure to test a hypothesis.** Think back to the lecture on whether you really wanted to test a hypothesis. If you are doing a power analysis but realize you have no idea what the standard deviation is, maybe you first need to spend more time validating the measures you are using.

When performing simulations or power analyses the same cautionary note can be made for estimates of correlations between dependent variables. When you are estimating these values from a sample, and want to perform simulations and/or power analyses, be aware that all estimates have some uncertainty. Try to get estimates which are as accurate as possible from pre-existing data. If possible, be a bit more conservative in sample size calculations based on estimated parameters, just to be sure.