Exploration 3.3: Body temperatures

2SD and Theory-Based Confidence Intervals for a Single Mean

LEARNING GOALS

- Approximate a 95% confidence interval for a mean by using the 2SD method.
- Compute a confidence interval for a mean using the theory-Based approach, including checking validity conditions.

In this section, we will focus on data consisting of a single quantitative variable. Hence, we will make inferences about a population mean by creating confidence intervals. We will again consider a shortcut 2SD approach as well as a theory-based approach.

Body temperatures? (cont.)

Even though 98.6 is often cited as the normal body temperature of a healthy person, many believe it is much lower than that. In his 1868 book, German Physician Carl August Wunderlich established what has been since assumed to be the average healthy human body temperature of 37° C or 98.6° F. In recent years, that number was thought to be too high. Protsiv et al. (2020) explored three large data sets that included body temperature, one collected between 1892 and 1930, one collected between 1971 and 1975, and one collected between 2007 and 2017. They suggested that not only is 98.6° F too large to be the average, but body temperatures have been decreasing steadily over time. Let's test this idea using students at your school as the subjects.

1. Let μ represent the average body temperature of all the students at your school. The research question is, "Is the mean body temperature for students at your school less than 98.6^{o}F?" If you still have the data from exploration 2.3, give the summary statistics here. If not, collect a new set of data from your class and give those summary statistics here.

Sample size, n =
Sample mean, $\bar{x}$ =
Sample SD, s =

We could conduct a simulation to generate a sampling distribution of the sample mean body temperature, assuming different values for μ to determine which are plausible. (In the **One Mean** applet, you can use a slider below the population to quickly change the value of the population mean.) Recall that such an interval is called a confidence interval. But this would be rather tedious, so in this exploration we will consider some shortcuts to estimating a confidence interval for μ.

2. Based on your sample data, what is your best guess or estimate for the value of μ, that is, the average body temperature for students at your school?

3. Suppose you repeated the survey by asking a different group of students at your school. Will the average for this new sample be exactly the same as the original sample average reported in #2? Why or why not?

A natural guess or estimate for the population average body temperature for students at your school (μ) is the observed value of the sample average body temperature. But, we know that if we were to conduct this study again we would most likely observe a different value for the sample average due to *sampling variability*—we need to account for this in our estimate for μ. To do so, we create a confidence interval for μ by using the formula

sample mean ± *multiplier* × (*SD of sample mean*)

where the "*SD of sample mean*" accounts for the sample-to-sample variability that we expect in the sample average, the "*multiplier*" accounts for the confidence level, and the "±" allows us to create an interval around the sample mean (these are the plausible parameter values that are not too far from the observed sample mean). Recall that when the confidence level is 95%, the multiplier is approximately 2. To find a 95% confidence interval for μ we can use the 2SD method, but we need to know what to use for the *SD of the sample mean*.

4. In #15 of Exploration 2.3, when you simulated the distribution of sample means from the hypothetical population of body temperatures, what did you find for the standard deviation of the sampling distribution? How did this value compare to the standard deviation of your sample?

Key Idea
There will be less sample-to-sample variability in the sample means than person-t-person variability in body temperatures.

5. In Exploration 2.2, we saw that $s/\sqrt{n}$ could be used to approximate the standard deviation of the sample means. We often refer to such an estimate of the standard deviation of a statistic as the *standard error*. Report this value.

Key Idea
An approximate 2SD method for a 95% confidence interval for a population mean, μ, is $\bar{x}$ ± 2 x $s/\sqrt{n}$
This method is valid when the sampling distribution follows a bell-shaped, symmetric distribution.

6. Use the value of $s/\sqrt{n}$ you found in #5 and the 2SD method to approximate a 95% confidence interval for μ in the context of the body temperature study, and then interpret this interval. (*Hint:* When interpreting the interval, be clear about what the interval estimates. You are estimating not the individual body temperatures of students but the body temperature of students at your school on ...)

7. Does the 95% confidence interval reported in #6 contain 98.6 (the population mean under the original null hypothesis)? Explain how you could have known this in advance based on your p-value from Exploration 2.3.

Theory-Based Approach

In Section 2.3, you saw a theory-based approach to find the p-value when testing hypotheses about a single population mean. This approach uses a more precise value of the multiplier in the confidence interval formula based on the *t*-distribution.

8. Go to the **Theory-Based Inference** applet, and select the **One Mean** option from the **Scenario** pull-down menu. Then, you can either copy and paste your class data or just enter the sample statistics (as recorded in #1). Press **Calculate**, check the **Confidence Interval** box, and press **Calculate CI**. Report the 95% confidence interval and compare it to the interval you found in #6. Are they about the same? Same midpoint? Same width? Could you have anticipated this? Explain your reasoning.

Recall that the theory-based approach uses a theoretical probability distribution called the *t*-distribution to predict the behavior of the standardized statistic (denoted by *t*) and that this approach required certain validity conditions to be met.

Validity Conditions

The theory-based interval for a population mean (called a ***one-sample t-interval***) requires that the quantitative variable should have a symmetric distribution or you should have at least 20 observations and the sample distribution should not be strongly skewed.

9. Are the validity conditions for using the 2SD or *t*-distribution based method to find a confidence interval met for your class data? Explain why or why not.

In summary, the formula for a theory-based confidence interval for a population mean (μ) is given by

$\bar{x}$ ± 2 x $s/\sqrt{n}$

where $\bar{x}$ denotes the sample mean, *s* denotes the sample standard deviation, and *n* is the sample size. We will rely on technology to find the appropriate multiplier for our sample size and our confidence level. Keep in mind that the multiplier is approximately 2 for 95% confidence and increases if we increase the confidence level.

Protsiv M, Ley C, Lankester J, Hastie T, Parsonnet J, "Decreasing human body temperature in the United States since the Industrial Revolution," eLife 2020;9:e49555, https://elifesciences.org/articles/49555

MIGHT NEED TO USE THIS

Take your body temperature and combine yours with those of your classmates. Alternately, you can use the dataset **ClassBodyTemperature**. The class data provided is given in degrees Fahrenheit. If you are collecting your own, it can be done in either Fahrenheit or Celsius. You just need to be consistent. If you use Celsius, just use 37° C for the average body temperature instead of 98.6° F.