Exploration 2.3: Have human body temperatures been decreasing?

Inference for a population mean

LEARNING GOALS

- Conduct a simulation-based analysis to conduct a single test involving the mean of a single quantitative variable.
- Anticipate the relative magnitude of the standard deviation of a sampling distribution of sample means based on the formula
- Carry out a theory-based analysis (one-sample *t*-test) involving the mean of a single quantitative variable, including checking relevant validity conditions.

In his 1868 book, German Physician Carl August Wunderlich established what has since been 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.

STEP 1: Ask a research question.

1. Is the mean body temperature for students at your school less than 98.6° F (what at one time was thought to be average)? Based on this question:

    a. What is the population of interest?

    b. What is the parameter?

    c. What symbol do we use for this parameter?

STEP 2: Design a study and collect data.

2. Write out the null and alternative hypotheses in words or in symbols for investigating the research question.

3. To test these hypotheses we need to collect data. Ideally, how should you obtain a sample from the population of all students at your school?

4. What variable would you measure on each student?

Obtaining a simple random sample of students from your school is an ideal way to attempt to get a representative sample. However, there is no question that it would take a fair bit of work to make that happen. This happens frequently in Statistics, and so researchers often opt to use a convenience sample instead. Researchers in the study mentioned earlier did not use data that came from simple random samples. One sample came from union army veterans of the American Civil War. These subjects were clearly not randomly chosen, and one could make a good argument they are not representative of the U.S. population at that time. However, with this sample they were able to compare a fairly wide age range of men taken over many years. The second sample came from the National Health and Nutrition Examination Survey. Again these subjects weren't randomly selected. However, they were chosen in such a way so that the sample was thought to be fairly representative of the U.S. population. The third dataset came from the Stanford Translational Research Integrated Database Environment. This project collects data from patients at Stanford University Medical Center. Again, clearly not a random sample of Americans. However, the authors felt that the last two samples were representative (after some adjustments were made) of all Americans. In the case of the Civil War veterans, they could compare temperatures of these males over time and with males from other samples.

5. One choice of a convenience sample of students at your school is to use your class. There are probably reasons why using your class as a sample could be representative of all students at your school in terms of body temperature. There are also probably reasons why it wouldn't be good. Identify at least one reason why your class might be a good representation of all students at your school in terms of body temperature and one reason why it wouldn't. (*Hint*: Body temperatures vary throughout the day typically lowest early in the morning and highest in the evening. Females tend to have higher body temperatures than males. Temperatures tend to increase after eating or exercising. Of course, illness also increases body temperature.)

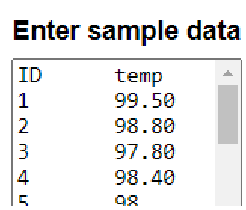
6. Fill in the blanks.
   Recall that convenience sampling may be __________ (biased/unbiased) whereas simple random sampling is __________ (biased/unbiased).

We will revisit the implications of using your class as a sample of students from your school later when we draw our final conclusions from the study. For now, let's gather the data. 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.

Key Idea
When describing a quantitative variable, you should always be aware of the measurement units of the variable. Measurement units indicate information about the context and scaling of the variable (e.g., hours vs. days).

STEP 3: Explore the data.
After you collect data, the next step is to explore the data. In the case of a quantitative variable (like we have here), exploring the data is a bit more involved than for a binary categorical variable. We can summarize a quantitative variable using a dotplot and then describe the shape, center, variability, and unusual observations.

Use the **One variable sampling** applet to examine a dotplot of the sample data. To do this, press **Clear** to erase the existing data and paste in your class data into the Data window. Then press **Use Data**.



7. Summarize the shape, center (mean with appropriate symbol and measurement units), and variability (standard deviation with appropriate symbol and measurement units) for the sample of body temperatures for your class. Also be sure to discuss any unusual observations or outliers in the data.

8. Do these data provide any preliminary evidence that students at your school *tend to* have body temperatures less than 98.6° F? Explain.

STEP 4: Draw inference beyond the data.
Let's now explore what the distribution of this sample tells us about the strength of evidence against the null hypothesis.

9. Suppose we had taken many, many samples of the same size from the population at your school. If the null hypothesis is true, how do you predict the distribution of sample means will behave?

Simulation-based Approach: Applying the 3S Strategy to Hypotheses about a Population Mean

To evaluate the strength of evidence that your sample data provide against the null hypothesis we need to investigate how often a sample mean as small as the one observed in your sample would occur if you were selecting a random sample from a population in which the mean body temperature really is 98.6° F (null hypothesis). In essence, we will apply the 3S strategy as we have done before.

We have the statistic (your sample mean from #7). Now we need to conduct a simulation. To start doing this, let's first convert your data set into something similar to a population.

- Assuming the population at your school is about 100 times larger than your class, we'll first make 100 copies of your sample data. To do this, just click on the **×100** button in the applet.
- Now we will model the null hypothesis by sliding your population distribution so that it centers on 98.6° F. To do this, click on the slider below the population in the applet and slide it so the population mean is as close to 98.6° F as you can get it. (*Hint:* Once you click on the slider you can use the arrow keys on your computer to slide it.)

10. What is the value of the standard deviation of your "population?" Use the symbol σ to refer to this value. (*Note:* You might think the population standard deviation will be the same value as your sample standard deviation since you just made many copies of the data, shifted it, but didn't change its variability. However, this difference occurs because population standard deviations are calculated slightly differently than sample standard deviations.)

Now let's take samples from this hypothetical population (assuming the null hypothesis is true) to see what a sampling distribution of sample means might look like. To do this check the **Show Sampling Options** box. Set the **Number of samples** to 10,000 and set the **Sample size** to match the sample size of your class data. Press **Draw Samples**.

11. Is the distribution of sample means centered around the hypothetical population mean? Is the distribution of sample means approximately normal? (Why?) Is the standard deviation of sample means close to $\sigma/\sqrt{n}$? (Why?)

12. Now for the strength of evidence. Put the observed sample mean (from #7) in the **Count Samples** box, choose direction based on the alternative hypothesis and find a p-value. Do you have strong evidence that the mean body temperature of students at your school is less than 98.6° F?

Theory-based Approach

You could overlay a normal distribution to get a theory-based p-value that should be similar to your simulation-based p-value. However, there is a slight problem with this. This normal distribution is constructed assuming the population standard deviation is known. Our population standard deviation was taken from our sample. In reality, the population probably has a standard deviation that is a bit different. Because of this, a slightly different distribution is used to predict the behavior of the standardized statistics. This new distribution is called a ***t*-distribution**. The *t*-distribution looks a lot like a normal distribution, but the shape is a bit flatter (more observations in the "tails," fewer in the middle). These "heavier tails" work for us here, where we have sample to sample variation in the means *plus* additional uncertainly by using the sample standard deviation each time rather than one population standard deviation. For this reason, the theory-based p-value for testing a sample mean will be based on the *t*-distribution rather than a normal distribution.

> **Key Idea**
>
> We can calculate a standardized statistic based on the observed sample data by computing
>
> $$\text{standardized statistic } = t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$
>
> (where μ_0 is a symbol used to represent the hypothesized value of the population mean).

13. Use the formula in the Key Idea box to calculate the *t*-statistic using the observed sample standard deviation.

14. Just like with standardized statistic when a single proportion is the statistic, 2 or -2 can be used for a reasonable cut-off value for declaring a *t*-statistic to be out in the tail of the null distribution. Is your calculated *t*-statistic less than -2? Would you consider your sample mean unlikely under a true null based on your *t*-statistic?

15. Now let's convert your distribution of sample means to a distribution of *t*-statistics. To do this, click on the *t*-statistic radio button above the null distribution in the applet.

    a. Use this distribution of simulated *t*-statistics to estimate the p-value by putting in the t-statistic you calculated in #13.

    b. The beauty of using the *t*-statistic is that their distribution can be predicted using a theoretical *t*-distribution. To do this, check **Overlay *t* Distribution** box. What is the value of the theory-based p-value given in the applet? How does it compare with your simulation-based p-values?

    c. Using your theory-based p-value, does your sample data provide strong evidence that the population mean body temperature for students at your school is less than 98.6° F? Justify your answer, as if you were talking to a friend who has not studied statistics.

As with a single proportion, theory-based tests are only valid when certain conditions are met. In the case of a test on a single mean (also called a **one-sample *t*-test**) the validity conditions are a little more complicated than they were for a single proportion.

Validity Conditions for a One-Sample* t*-test

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 for a p-value from a one-sample *t*-test to be valid.

16. Does your temperature data meet the criteria for a valid one-sample *t*-test? Explain.

STEP 5: Formulate conclusions.

The researchers exploring the decline in average body temperature over time discussed a number of factors that might explain this decrease. They thought that a reduction in inflammation in humans over the last century and a half was the most plausible explanation for this observed decrease. Many diseases and conditions that were once common have been irradicated or greatly controlled through the development of vaccinations and other treatments. Therefore, humans have been getting healthier and living longer. Other reasons for lower body temperatures could be the ability to maintain a more constant body temperature through heating and air conditioning and the loss of microbial diversity and increased antibiotic use.

17. Think again about how your sample was selected from the population. Do you feel comfortable generalizing the results of your analysis to all students at your school? Explain.

STEP 6: Look back and ahead.

There have been different methods of obtaining body temperature over time and different types of thermometers to do this. This was a concern of the researchers exploring the decline in average body temperature. They did not think this was a factor, however, because when the same method of obtaining temperatures was used, they still saw a decrease in based on birth year. They also saw this consistently within all subgroups that they looked at.

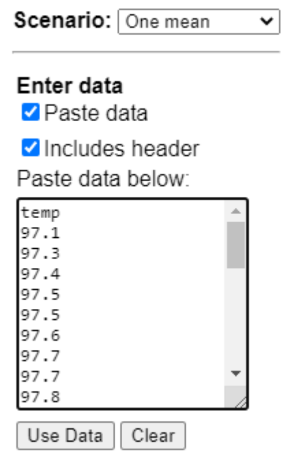
18. Did anything about the design of your study limit you in the conclusions you made? Issues you may want to critique include:
    - The match between the research question and the study design
    - How the observational units were selected
    - How the measurements were recorded

19. Looking ahead: What might be some next steps you would take to fix the limitations you found or build on your conclusions?

Exploring Further

According to WebMD https://www.webmd.com/first-aid/normal-body-temperature, "… more recent studies say the baseline (average body temperature) for most people is closer to 98.2° F." Use your data and the **Theory-Based Inference** applet to test to see whether your sample provides strong evidence that average body temperature for students at your school is different than 98.2° F.

To run this test, put your data in the [Theory-Based Inference] applet. To do this, change the **Scenario** to One mean, click on the **Paste data** box, and the **Includes header** box. Press **Clear** to delete the existing data in the applet and then copy and paste your class data (with a one-word variable name) into the **Paste data below** box and press **Use Data**. To run the test, click on the **Test of significance** box, put in the appropriate numbers for the hypotheses, chose the appropriate sign for the alternative hypothesis, and click **Calculate**.



20. Give your hypotheses in symbols, p-value and conclusion for the test to see whether your sample provides strong evidence that average body temperature for students at your school is *different* than 98.2° F.

21. Suppose you did not find strong evidence strong evidence that average body temperature for students at your school is different than 98.2° F in #20.

    a. Does that mean you have strong evidence that the population mean is 98.2° F? Explain.

    b. Does that mean you have *proven* that the population mean is 98.2° F? Explain.

Reference:

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