

## 7.2: The Student's T-Distribution

In a previous lesson, we saw that under certain conditions, the sampling distribution of sample means is normally distributed. We can use this information to standardize the sampling distribution of sample means and find probabilities and Z-scores from the standard normal distribution. We can do this if we know the population standard deviation. However, it is not common to know a population standard deviation. What happens when we don't know the value of the population standard deviation?

### When the Population Standard Deviation, $\sigma$ , is Unknown

In most cases, we do not know the population standard deviation,  $\sigma$ . The only option we have is to approximate with a known sample standard deviation,  $s$ .

1. Review: What conditions should be met to verify that the sampling distribution of sample means is approximately normal?

Recall, the mean of the sampling distribution of sample means is  $\mu_{\bar{x}} = \mu$ . The standard error of the sampling distribution of sample means is  $\mu_{\bar{x}} = \mu$ . We substitute  $s$  for in the standard error formula since is unknown. So, the standard error of the sampling distribution of sample means is estimated by

$$\frac{s}{\sqrt{n}}$$

The test statistic for a sample mean is  $\frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$ . The sample standard deviation,  $s$ , varies from sample to sample and it only

approximates  $\sigma$ . Therefore, this substitution introduces additional variability into the test statistic. This added variability means that the distribution of the test statistic is no longer normal. We instead use the Student's T-distribution.

### The T-Distribution

The **T-distribution** describes the variability of the test statistic,  $T = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$ , when the sampling distribution of sample means is

normal, and the sample standard deviation,  $s$ , is used to estimate an unknown population standard deviation,  $\sigma$ .

This test statistic is an estimate of how many standard errors the sample mean is from the hypothesized population mean.

A T-distribution is a member of a family of continuous probability distributions. The width of a T-distribution depends on how much a sample standard deviation can vary. The amount of variability in a sample standard deviation depends on how many deviations *vary freely* when it is computed. Recall, the sample standard deviation is calculated using the formula

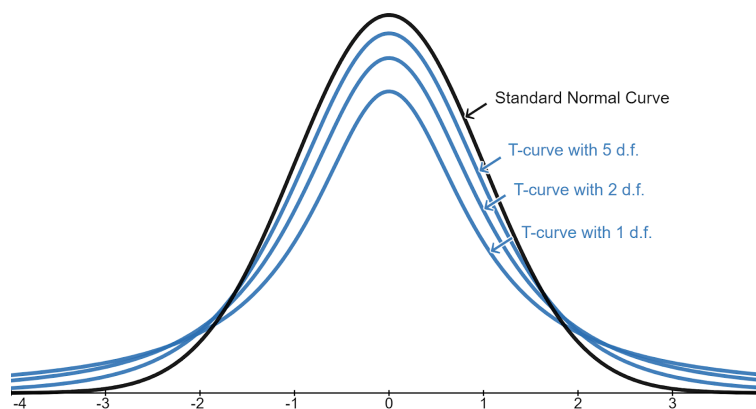
$$s = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n - 1}}$$

The deviations from the mean are averaged in a sample standard deviation. When the deviations are added together, they always add to zero. Because of this, the last deviation summed in this average is not free – it is always the value that makes the resulting sum zero. There are  $n$  deviations from the mean, but only  $n - 1$  of these are *free* deviations.

The variability of standard deviations depends on the number of free deviations in the sample standard deviations,  $n - 1$ . This quantity is known as the **degrees of freedom (d.f.)**. Each degree of freedom defines a uniquely associated T-distribution.

T-distributions have the following characteristics:

- T-distributions are bell-shaped and symmetric with a mean of 0.
- Each T-distribution depends on the degrees of freedom, d.f. ( $d.f. = n - 1$ ).
- T-distributions have heavier tails and narrower peaks than the standard normal distribution.
- The total area under each T-distribution curve is 1.
- As the degrees of freedom increase, the tails become thinner and the curve approaches the standard normal distribution.



Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

With fewer degrees of freedom, the more the sample standard deviation varies. In other words, a smaller sample size corresponds to greater variability in a T-distribution. In a small sample, it is more likely to observe extreme values. This is reflected in the shape of the T-distribution. As the sample size increases, the T-distribution trends toward the normal distribution.

Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

### Check for Understanding

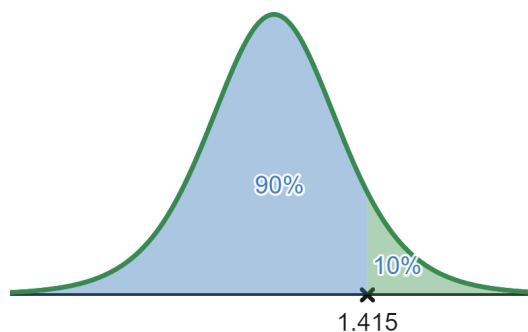
2. The heights of students at Claremont High School is approximately normal. If the average height of a student here is 166 centimeters.
  - a. Is the sampling distribution of sample means better represented by the normal distribution or the T-distribution? Justify your answer.
  - b. You find the average height and sample standard deviation for 10 randomly selected Claremont High School students. The sample standard deviation is 4 cm. Compute the test statistic for an average height of 168 cm. Round your answer to two decimal places.

- c. We will now use [desmos](https://www.desmos.com/calculator) to find the likelihood of seeing an average height or 168 cm or more.
- Go to <https://www.desmos.com/calculator>
  - The degrees of freedom d.f. in this example is  $n - 1 = 10 - 1 = 9$ . Type `tdist(9)`. In general, type `tdist(degrees of freedom)`.
  - Click the “Zoom Fit” button.
  - Check the “Find Cumulative Probability (CDF)” box.
  - The min and max will default to  $-\infty$  and  $\infty$  respectively. Enter the test statistics from b. into the min field. The probability  $P(\bar{x} \geq 168) = P(T \geq \text{---}) \approx 0.0743$ .
- d. You find the average height and sample standard deviation for 10 randomly selected Claremont High School students. The sample standard deviation is 4 cm. Use the process in c. to compute the probability of observing an average height of 170 cm or more.
- e. If the average height of a student at Claremont High School is 166 centimeters, which of the following is more likely:
- a random sample of 15 students, with the average height being greater than 170 centimeters? OR
  - a random sample of 25 students, with the average height being greater than 170 centimeters?

Justify your answer using what you know about the T-distribution.

## Finding a Critical Value from a T-Distribution

3. We will use [desmos](#) to find the T critical value that separates the top 10% from the lower 90% from a random sample of size 8.
- Go to <https://www.desmos.com/calculator>
  - The degrees of freedom  $d.f. = n - 1 = 8 - 1 =$  \_\_\_\_\_. The area to the left of the critical value is \_\_\_\_\_ (90%). Type `tdist(7).inversecdf(0.9)`.
    - In general, you can type `tdist(d.f.).inversecdf(L)` to find a T critical value where d.f. is the degrees of freedom, and L is the area left of the critical value.



The T critical value here is around 1.415

Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

4. Use [desmos](#) to find the T critical value that separates the middle 95% from a random sample of size 8. Show your thinking by sketching a graph.

---

This page titled [7.2: The Student's T-Distribution](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Hannah Seidler-Wright](#).

- [Current page](#) by Hannah Seidler-Wright is licensed [CC BY-NC-SA 4.0](#).
- [1.2: The Statistical Analysis Process](#) by Hannah Seidler-Wright is licensed [CC BY-NC-SA 4.0](#).