



Section 7.1: One-Sample Means with the T-Distribution

Diez, D. M., Çetinkaya-Rundel, M., Barr, C. D. (2019). OpenIntro Statistics (4th ed.). OpenIntro.
<https://www.openintro.org/book/os/> CC BY-SA 3.0

STAT 1201
Introduction to Probability and Statistics

ONLINE AND DISTANCE EDUCATION

Chapter 7

Inference for numerical data

7.1 One-sample means with the t -distribution

7.2 Paired data

7.3 Difference of two means

7.4 Power calculations for a difference of means

7.5 Comparing many means with ANOVA

Chapter 5 introduced a framework for statistical inference based on confidence intervals and hypotheses using the normal distribution for sample proportions. In this chapter, we encounter several new point estimates and a couple new distributions. In each case, the inference ideas remain the same: determine which point estimate or test statistic is useful, identify an appropriate distribution for the point estimate or test statistic, and apply the ideas of inference.



For videos, slides, and other resources, please visit
www.openintro.org/os

7.1 One-sample means with the t -distribution

Similar to how we can model the behavior of the sample proportion \hat{p} using a normal distribution, the sample mean \bar{x} can also be modeled using a normal distribution when certain conditions are met. However, we'll soon learn that a new distribution, called the t -distribution, tends to be more useful when working with the sample mean. We'll first learn about this new distribution, then we'll use it to construct confidence intervals and conduct hypothesis tests for the mean.

7.1.1 The sampling distribution of \bar{x}

The sample mean tends to follow a normal distribution centered at the population mean, μ , when certain conditions are met. Additionally, we can compute a standard error for the sample mean using the population standard deviation σ and the sample size n .

CENTRAL LIMIT THEOREM FOR THE SAMPLE MEAN

When we collect a sufficiently large sample of n independent observations from a population with mean μ and standard deviation σ , the sampling distribution of \bar{x} will be nearly normal with

$$\text{Mean} = \mu \qquad \text{Standard Error (SE)} = \frac{\sigma}{\sqrt{n}}$$

Before diving into confidence intervals and hypothesis tests using \bar{x} , we first need to cover two topics:

- When we modeled \hat{p} using the normal distribution, certain conditions had to be satisfied. The conditions for working with \bar{x} are a little more complex, and we'll spend Section 7.1.2 discussing how to check conditions for inference.
- The standard error is dependent on the population standard deviation, σ . However, we rarely know σ , and instead we must estimate it. Because this estimation is itself imperfect, we use a new distribution called the t -distribution to fix this problem, which we discuss in Section 7.1.3.

7.1.2 Evaluating the two conditions required for modeling \bar{x}

Two conditions are required to apply the Central Limit Theorem for a sample mean \bar{x} :

Independence. The sample observations must be independent. The most common way to satisfy this condition is when the sample is a simple random sample from the population. If the data come from a random process, analogous to rolling a die, this would also satisfy the independence condition.

Normality. When a sample is small, we also require that the sample observations come from a normally distributed population. We can relax this condition more and more for larger and larger sample sizes. This condition is obviously vague, making it difficult to evaluate, so next we introduce a couple rules of thumb to make checking this condition easier.

RULES OF THUMB: HOW TO PERFORM THE NORMALITY CHECK

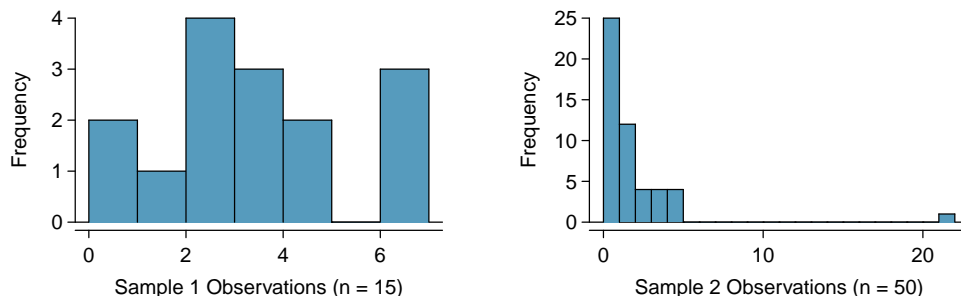
There is no perfect way to check the normality condition, so instead we use two rules of thumb:

- $n < 30$:** If the sample size n is less than 30 and there are no clear outliers in the data, then we typically assume the data come from a nearly normal distribution to satisfy the condition.
- $n \geq 30$:** If the sample size n is at least 30 and there are no *particularly extreme* outliers, then we typically assume the sampling distribution of \bar{x} is nearly normal, even if the underlying distribution of individual observations is not.

In this first course in statistics, you aren't expected to develop perfect judgement on the normality condition. However, you are expected to be able to handle clear cut cases based on the rules of thumb.¹

EXAMPLE 7.1

Consider the following two plots that come from simple random samples from different populations. Their sample sizes are $n_1 = 15$ and $n_2 = 50$.



E

Are the independence and normality conditions met in each case?

Each sample is from a simple random sample of its respective population, so the independence condition is satisfied. Let's next check the normality condition for each using the rule of thumb.

The first sample has fewer than 30 observations, so we are watching for any clear outliers. None are present; while there is a small gap in the histogram on the right, this gap is small and 20% of the observations in this small sample are represented in that far right bar of the histogram, so we can hardly call these clear outliers. With no clear outliers, the normality condition is reasonably met.

The second sample has a sample size greater than 30 and includes an outlier that appears to be roughly 5 times further from the center of the distribution than the next furthest observation. This is an example of a particularly extreme outlier, so the normality condition would not be satisfied.

In practice, it's typical to also do a mental check to evaluate whether we have reason to believe the underlying population would have moderate skew (if $n < 30$) or have particularly extreme outliers ($n \geq 30$) beyond what we observe in the data. For example, consider the number of followers for each individual account on Twitter, and then imagine this distribution. The large majority of accounts have built up a couple thousand followers or fewer, while a relatively tiny fraction have amassed tens of millions of followers, meaning the distribution is extremely skewed. When we know the data come from such an extremely skewed distribution, it takes some effort to understand what sample size is large enough for the normality condition to be satisfied.

7.1.3 Introducing the t -distribution

In practice, we cannot directly calculate the standard error for \bar{x} since we do not know the population standard deviation, σ . We encountered a similar issue when computing the standard error for a sample proportion, which relied on the population proportion, p . Our solution in the proportion context was to use sample value in place of the population value when computing the standard error. We'll employ a similar strategy for computing the standard error of \bar{x} , using the sample standard deviation s in place of σ :

$$SE = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$$

This strategy tends to work well when we have a lot of data and can estimate σ using s accurately. However, the estimate is less precise with smaller samples, and this leads to problems when using the normal distribution to model \bar{x} .

¹More nuanced guidelines would consider further relaxing the *particularly extreme outlier* check when the sample size is very large. However, we'll leave further discussion here to a future course.

We'll find it useful to use a new distribution for inference calculations called the **t -distribution**. A t -distribution, shown as a solid line in Figure 7.1, has a bell shape. However, its tails are thicker than the normal distribution's, meaning observations are more likely to fall beyond two standard deviations from the mean than under the normal distribution. The extra thick tails of the t -distribution are exactly the correction needed to resolve the problem of using s in place of σ in the SE calculation.

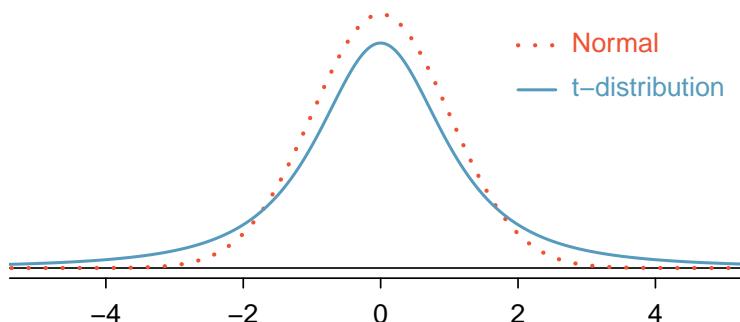


Figure 7.1: Comparison of a t -distribution and a normal distribution.

The t -distribution is always centered at zero and has a single parameter: degrees of freedom. The **degrees of freedom (df)** describes the precise form of the bell-shaped t -distribution. Several t -distributions are shown in Figure 7.2 in comparison to the normal distribution.

In general, we'll use a t -distribution with $df = n - 1$ to model the sample mean when the sample size is n . That is, when we have more observations, the degrees of freedom will be larger and the t -distribution will look more like the standard normal distribution; when the degrees of freedom is about 30 or more, the t -distribution is nearly indistinguishable from the normal distribution.

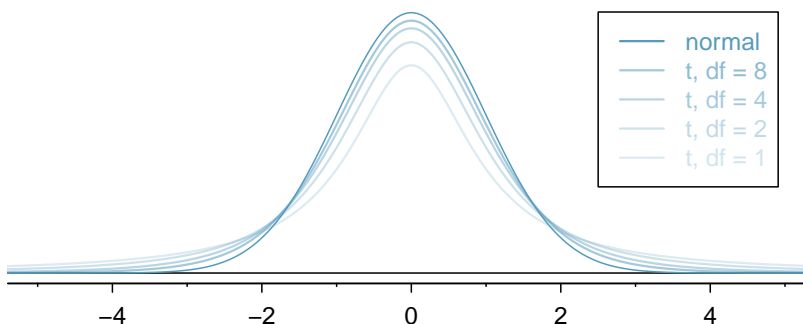


Figure 7.2: The larger the degrees of freedom, the more closely the t -distribution resembles the standard normal distribution.

DEGREES OF FREEDOM (df)

The degrees of freedom describes the shape of the t -distribution. The larger the degrees of freedom, the more closely the distribution approximates the normal model.

When modeling \bar{x} using the t -distribution, use $df = n - 1$.

The t -distribution allows us greater flexibility than the normal distribution when analyzing numerical data. In practice, it's common to use statistical software, such as R, Python, or SAS for these analyses. Alternatively, a graphing calculator or a **t -table** may be used; the t -table is similar to the normal distribution table, and it may be found in Appendix C.2, which includes usage instructions and examples for those who wish to use this option. No matter the approach you choose, apply your method using the examples below to confirm your working understanding of the t -distribution.

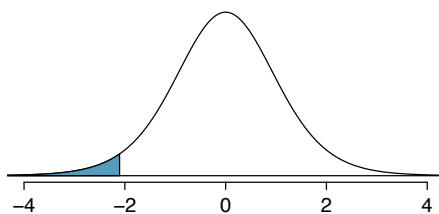


Figure 7.3: The t -distribution with 18 degrees of freedom. The area below -2.10 has been shaded.

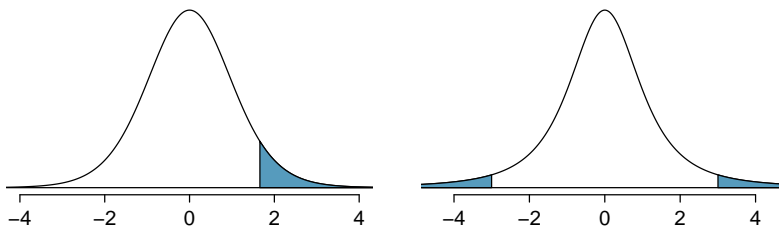


Figure 7.4: Left: The t -distribution with 20 degrees of freedom, with the area above 1.65 shaded. Right: The t -distribution with 2 degrees of freedom, with the area further than 3 units from 0 shaded.

EXAMPLE 7.2

What proportion of the t -distribution with 18 degrees of freedom falls below -2.10 ?

E

Just like a normal probability problem, we first draw the picture in Figure 7.3 and shade the area below -2.10 . Using statistical software, we can obtain a precise value: 0.0250 .

EXAMPLE 7.3

A t -distribution with 20 degrees of freedom is shown in the left panel of Figure 7.4. Estimate the proportion of the distribution falling above 1.65 .

E

With a normal distribution, this would correspond to about 0.05 , so we should expect the t -distribution to give us a value in this neighborhood. Using statistical software: 0.0573 .

EXAMPLE 7.4

A t -distribution with 2 degrees of freedom is shown in the right panel of Figure 7.4. Estimate the proportion of the distribution falling more than 3 units from the mean (above or below).

E

With so few degrees of freedom, the t -distribution will give a more notably different value than the normal distribution. Under a normal distribution, the area would be about 0.003 using the 68-95-99.7 rule. For a t -distribution with $df = 2$, the area in both tails beyond 3 units totals 0.0955 . This area is dramatically different than what we obtain from the normal distribution.

GUIDED PRACTICE 7.5

G

What proportion of the t -distribution with 19 degrees of freedom falls above -1.79 units? Use your preferred method for finding tail areas.²

²We want to find the shaded area *above* -1.79 (we leave the picture to you). The lower tail area has an area of 0.0447 , so the upper area would have an area of $1 - 0.0447 = 0.9553$.

7.1.4 One sample t -confidence intervals

Let's get our first taste of applying the t -distribution in the context of an example about the mercury content of dolphin muscle. Elevated mercury concentrations are an important problem for both dolphins and other animals, like humans, who occasionally eat them.



Figure 7.5: A Risso's dolphin.

Photo by Mike Baird (www.bairdphotos.com). CC BY 2.0 license.

We will identify a confidence interval for the average mercury content in dolphin muscle using a sample of 19 Risso's dolphins from the Taiji area in Japan. The data are summarized in Figure 7.6. The minimum and maximum observed values can be used to evaluate whether or not there are clear outliers.

n	\bar{x}	s	minimum	maximum
19	4.4	2.3	1.7	9.2

Figure 7.6: Summary of mercury content in the muscle of 19 Risso's dolphins from the Taiji area. Measurements are in micrograms of mercury per wet gram of muscle ($\mu\text{g}/\text{wet g}$).

EXAMPLE 7.6

Are the independence and normality conditions satisfied for this data set?

E

The observations are a simple random sample, therefore independence is reasonable. The summary statistics in Figure 7.6 do not suggest any clear outliers, since all observations are within 2.5 standard deviations of the mean. Based on this evidence, the normality condition seems reasonable.

In the normal model, we used z^* and the standard error to determine the width of a confidence interval. We revise the confidence interval formula slightly when using the t -distribution:

$$\text{point estimate} \pm t_{df}^* \times SE \quad \rightarrow \quad \bar{x} \pm t_{df}^* \times \frac{s}{\sqrt{n}}$$

EXAMPLE 7.7

Using the summary statistics in Figure 7.6, compute the standard error for the average mercury content in the $n = 19$ dolphins.

E

We plug in s and n into the formula: $SE = s/\sqrt{n} = 2.3/\sqrt{19} = 0.528$.

The value t_{df}^* is a cutoff we obtain based on the confidence level and the t -distribution with df degrees of freedom. That cutoff is found in the same way as with a normal distribution: we find t_{df}^* such that the fraction of the t -distribution with df degrees of freedom within a distance t_{df}^* of 0 matches the confidence level of interest.

EXAMPLE 7.8

When $n = 19$, what is the appropriate degrees of freedom? Find t_{df}^* for this degrees of freedom and the confidence level of 95%

E

The degrees of freedom is easy to calculate: $df = n - 1 = 18$.

Using statistical software, we find the cutoff where the upper tail is equal to 2.5%: $t_{18}^* = 2.10$. The area below -2.10 will also be equal to 2.5%. That is, 95% of the t -distribution with $df = 18$ lies within 2.10 units of 0.

EXAMPLE 7.9

Compute and interpret the 95% confidence interval for the average mercury content in Risso's dolphins.

E

We can construct the confidence interval as

$$\bar{x} \pm t_{18}^* \times SE \rightarrow 4.4 \pm 2.10 \times 0.528 \rightarrow (3.29, 5.51)$$

We are 95% confident the average mercury content of muscles in Risso's dolphins is between 3.29 and 5.51 $\mu\text{g}/\text{wet gram}$, which is considered extremely high.

FINDING A t -CONFIDENCE INTERVAL FOR THE MEAN

Based on a sample of n independent and nearly normal observations, a confidence interval for the population mean is

$$\text{point estimate} \pm t_{df}^* \times SE \rightarrow \bar{x} \pm t_{df}^* \times \frac{s}{\sqrt{n}}$$

where \bar{x} is the sample mean, t_{df}^* corresponds to the confidence level and degrees of freedom df , and SE is the standard error as estimated by the sample.

GUIDED PRACTICE 7.10

G

The FDA's webpage provides some data on mercury content of fish. Based on a sample of 15 croaker white fish (Pacific), a sample mean and standard deviation were computed as 0.287 and 0.069 ppm (parts per million), respectively. The 15 observations ranged from 0.18 to 0.41 ppm. We will assume these observations are independent. Based on the summary statistics of the data, do you have any objections to the normality condition of the individual observations?³

EXAMPLE 7.11

Estimate the standard error of $\bar{x} = 0.287$ ppm using the data summaries in Guided Practice 7.10. If we are to use the t -distribution to create a 90% confidence interval for the actual mean of the mercury content, identify the degrees of freedom and t_{df}^* .

E

The standard error: $SE = \frac{0.069}{\sqrt{15}} = 0.0178$.

Degrees of freedom: $df = n - 1 = 14$.

Since the goal is a 90% confidence interval, we choose t_{14}^* so that the two-tail area is 0.1: $t_{14}^* = 1.76$.

³The sample size is under 30, so we check for obvious outliers: since all observations are within 2 standard deviations of the mean, there are no such clear outliers.

CONFIDENCE INTERVAL FOR A SINGLE MEAN

Once you've determined a one-mean confidence interval would be helpful for an application, there are four steps to constructing the interval:

Prepare. Identify \bar{x} , s , n , and determine what confidence level you wish to use.

Check. Verify the conditions to ensure \bar{x} is nearly normal.

Calculate. If the conditions hold, compute SE , find t_{df}^* , and construct the interval.

Conclude. Interpret the confidence interval in the context of the problem.

GUIDED PRACTICE 7.12

- Ⓒ Using the information and results of Guided Practice 7.10 and Example 7.11, compute a 90% confidence interval for the average mercury content of croaker white fish (Pacific).⁴

GUIDED PRACTICE 7.13

- Ⓒ The 90% confidence interval from Guided Practice 7.12 is 0.256 ppm to 0.318 ppm. Can we say that 90% of croaker white fish (Pacific) have mercury levels between 0.256 and 0.318 ppm?⁵

7.1.5 One sample t -tests

Is the typical US runner getting faster or slower over time? We consider this question in the context of the Cherry Blossom Race, which is a 10-mile race in Washington, DC each spring.

The average time for all runners who finished the Cherry Blossom Race in 2006 was 93.29 minutes (93 minutes and about 17 seconds). We want to determine using data from 100 participants in the 2017 Cherry Blossom Race whether runners in this race are getting faster or slower, versus the other possibility that there has been no change.

GUIDED PRACTICE 7.14

- Ⓒ What are appropriate hypotheses for this context?⁶

GUIDED PRACTICE 7.15

- Ⓒ The data come from a simple random sample of all participants, so the observations are independent. However, should we be worried about the normality condition? See Figure 7.7 for a histogram of the differences and evaluate if we can move forward.⁷

When completing a hypothesis test for the one-sample mean, the process is nearly identical to completing a hypothesis test for a single proportion. First, we find the Z -score using the observed value, null value, and standard error; however, we call it a **T-score** since we use a t -distribution for calculating the tail area. Then we find the p-value using the same ideas we used previously: find the one-tail area under the sampling distribution, and double it.

⁴ $\bar{x} \pm t_{14}^* \times SE \rightarrow 0.287 \pm 1.76 \times 0.0178 \rightarrow (0.256, 0.318)$. We are 90% confident that the average mercury content of croaker white fish (Pacific) is between 0.256 and 0.318 ppm.

⁵ No, a confidence interval only provides a range of plausible values for a population parameter, in this case the population mean. It does not describe what we might observe for individual observations.

⁶ H_0 : The average 10-mile run time was the same for 2006 and 2017. $\mu = 93.29$ minutes. H_A : The average 10-mile run time for 2017 was *different* than that of 2006. $\mu \neq 93.29$ minutes.

⁷ With a sample of 100, we should only be concerned if there are particularly extreme outliers. The histogram of the data doesn't show any outliers of concern (and arguably, no outliers at all).

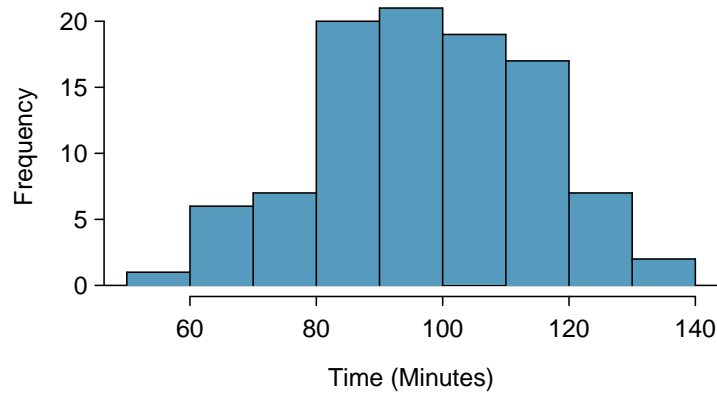


Figure 7.7: A histogram of `time` for the sample Cherry Blossom Race data.

EXAMPLE 7.16

With both the independence and normality conditions satisfied, we can proceed with a hypothesis test using the t -distribution. The sample mean and sample standard deviation of the sample of 100 runners from the 2017 Cherry Blossom Race are 97.32 and 16.98 minutes, respectively. Recall that the sample size is 100 and the average run time in 2006 was 93.29 minutes. Find the test statistic and p-value. What is your conclusion?

To find the test statistic (T-score), we first must determine the standard error:

$$SE = 16.98/\sqrt{100} = 1.70$$

Now we can compute the T -score using the sample mean (97.32), null value (93.29), and SE :

$$T = \frac{97.32 - 93.29}{1.70} = 2.37$$

For $df = 100 - 1 = 99$, we can determine using statistical software (or a t -table) that the one-tail area is 0.01, which we double to get the p-value: 0.02.

Because the p-value is smaller than 0.05, we reject the null hypothesis. That is, the data provide strong evidence that the average run time for the Cherry Blossom Run in 2017 is different than the 2006 average. Since the observed value is above the null value and we have rejected the null hypothesis, we would conclude that runners in the race were slower on average in 2017 than in 2006.

HYPOTHESIS TESTING FOR A SINGLE MEAN

Once you've determined a one-mean hypothesis test is the correct procedure, there are four steps to completing the test:

Prepare. Identify the parameter of interest, list out hypotheses, identify the significance level, and identify \bar{x} , s , and n .

Check. Verify conditions to ensure \bar{x} is nearly normal.

Calculate. If the conditions hold, compute SE , compute the T-score, and identify the p-value.

Conclude. Evaluate the hypothesis test by comparing the p-value to α , and provide a conclusion in the context of the problem.