One-sample means with the t-distribution

Colby Community College

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:

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

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:

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

Note

It's rare to need to estimate the population mean μ, but somehow know the population standard deviation σ. In most cases σ will need to be estimated.

Conditions to Apply the Central Limit Theorem

Independence: The sample observations must be independent.

Conditions to Apply the Central Limit Theorem

Independence: The sample observations must be independent.

Normality: When a sample is small, we also require that the sample observations come from a normally distributed population.

Conditions to Apply the Central Limit Theorem

Independence: The sample observations must be independent.

Normality: When a sample is small, we also require that the sample observations come from a normally distributed population.

- **$n < 30$:** If the sample size n is less than 30 and there are no clear outliers in the data, then we typically assume the data is from a nearly normal population.

Conditions to Apply the Central Limit Theorem

Independence: The sample observations must be independent.

Normality: When a sample is small, we also require that the sample observations come from a normally distributed population.

- $n < 30$: If the sample size n is less than 30 and there are no clear outliers in the data, then we typically assume the data is from a nearly normal population.
- $n \geq 30$: 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 population is not.

Conditions to Apply the Central Limit Theorem

Independence: The sample observations must be independent.

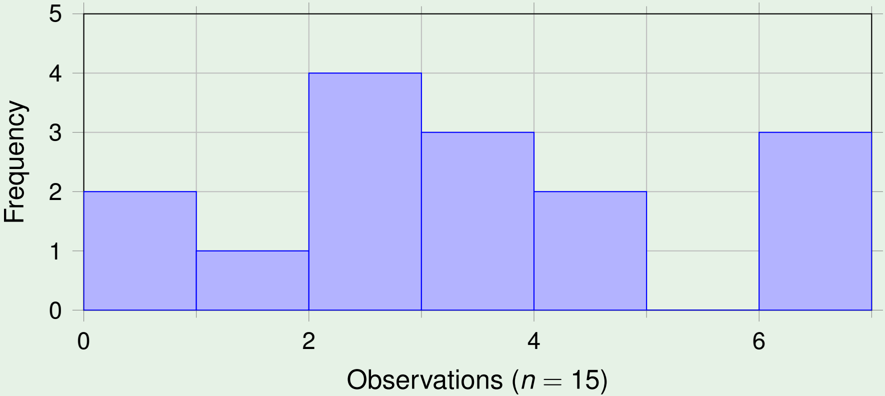
Normality: When a sample is small, we also require that the sample observations come from a normally distributed population.

- $n < 30$: If the sample size n is less than 30 and there are no clear outliers in the data, then we typically assume the data is from a nearly normal population.
- $n \geq 30$: 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 population is not.

Note

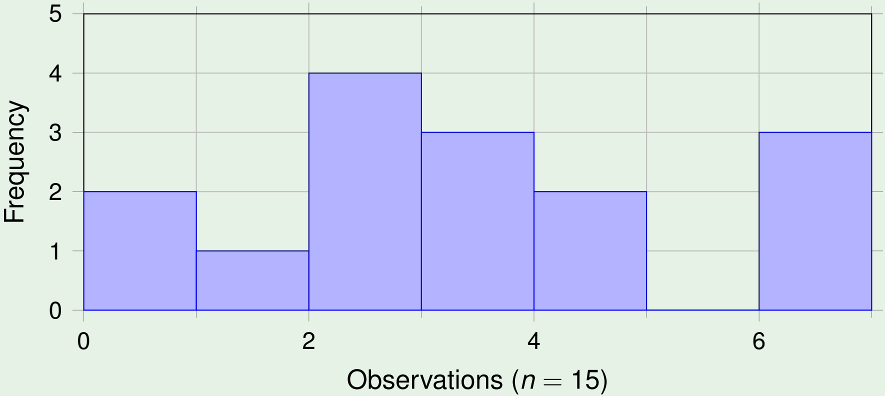
In a first course in statistics, you aren't expected to develop perfect judgment on the normality condition.

Example 1



Is the normality condition met?

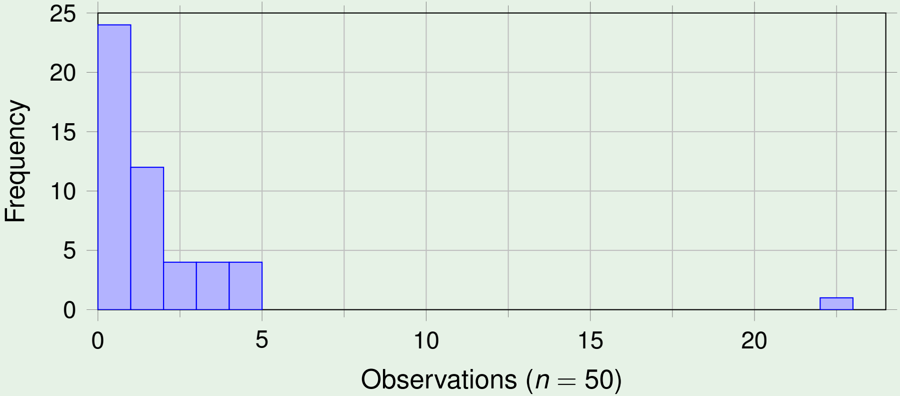
Example 1



Is the normality condition met?

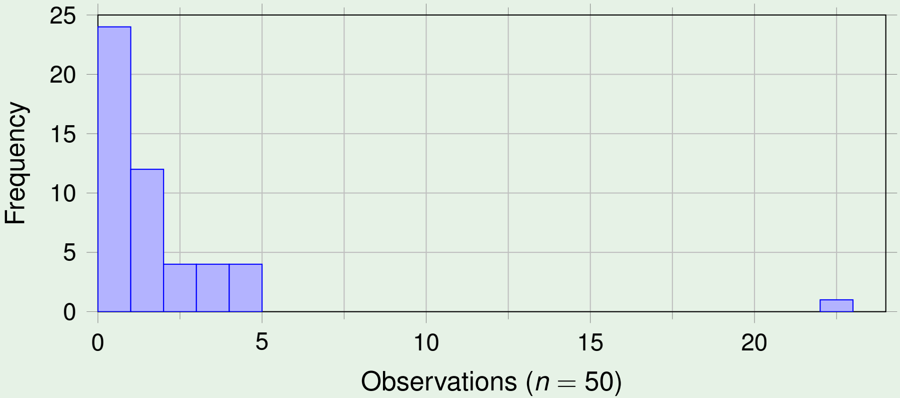
Since there are less than 30 observations, we need to look for *clear* outliers. While there is a gap on the right, the gap is small and 20% of the observations fall in rightmost bar. We can't really call these clear outliers, so the normality condition is reasonably met.

Example 2



Is the normality condition met?

Example 2



Is the normality condition met?

The sample size is greater than 30, so we need to look for an extreme outlier. The gap is more than four times the width of the cluster on the left side, so this is clearly an extreme outlier and the normality condition is not met.

Note

In practice, we cannot directly calculate the standard error for $\bar{x}$, since we do not know the population standard deviation σ.

Note

In practice, we cannot directly calculate the standard error for $\bar{x}$, since we do not know the population standard deviation σ.

We can use the sample standard deviation s as the best estimate of σ:

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

Note

In practice, we cannot directly calculate the standard error for $\bar{x}$, since we do not know the population standard deviation σ.

We can use the sample standard deviation s as the best estimate of σ:

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

Definition

If a population has a normal distribution, then the distribution of

$$t = \frac{\bar{x} - \mu}{SE} \approx \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

is called the **Student *t* distribution** for sample sizes n.

Note

In practice, we cannot directly calculate the standard error for $\bar{x}$, since we do not know the population standard deviation σ.

We can use the sample standard deviation s as the best estimate of σ:

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

Definition

If a population has a normal distribution, then the distribution of

$$t = \frac{\bar{x} - \mu}{SE} \approx \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

is called the **Student *t* distribution** for sample sizes n.

Note

A Student t distribution is commonly called a ***t* distribution**.

Definition

The **degrees of freedom** (or **df**) for a collection of sample data is the number of sample values that can vary after certain restrictions have been imposed on all data values.

When modeling $\bar{x}$ using the t-distribution, use:

$$df = n - 1$$

Definition

The **degrees of freedom** (or **df**) for a collection of sample data is the number of sample values that can vary after certain restrictions have been imposed on all data values.

When modeling $\bar{x}$ using the t-distribution, use:

$$df = n - 1$$

Example 3

If 10 test scores must have mean 80, then their sum must be 800.

Definition

The **degrees of freedom** (or **df**) for a collection of sample data is the number of sample values that can vary after certain restrictions have been imposed on all data values.

When modeling $\bar{x}$ using the *t*-distribution, use:

$$df = n - 1$$

Example 3

If 10 test scores must have mean 80, then their sum must be 800.

We can freely assign values to the first 9 scores, but the 10th score would need to be:

$$s_{10} = 800 - s_1 - s_2 - s_3 - s_4 - s_5 - s_6 - s_7 - s_8 - s_9$$

Definition

The **degrees of freedom** (or **df**) for a collection of sample data is the number of sample values that can vary after certain restrictions have been imposed on all data values.

When modeling $\bar{x}$ using the *t*-distribution, use:

$$df = n - 1$$

Example 3

If 10 test scores must have mean 80, then their sum must be 800.

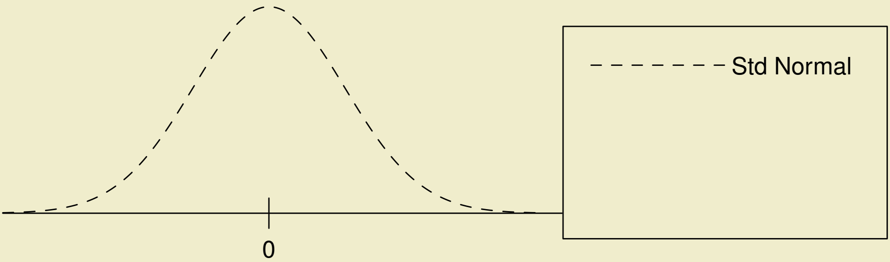
We can freely assign values to the first 9 scores, but the 10th score would need to be:

$$s_{10} = 800 - s_1 - s_2 - s_3 - s_4 - s_5 - s_6 - s_7 - s_8 - s_9$$

Hence 9 degrees of freedom.

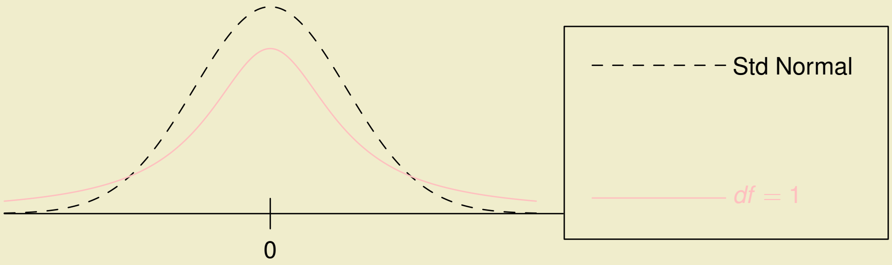
Note

The Student *t* distribution changes for different degrees of freedom.



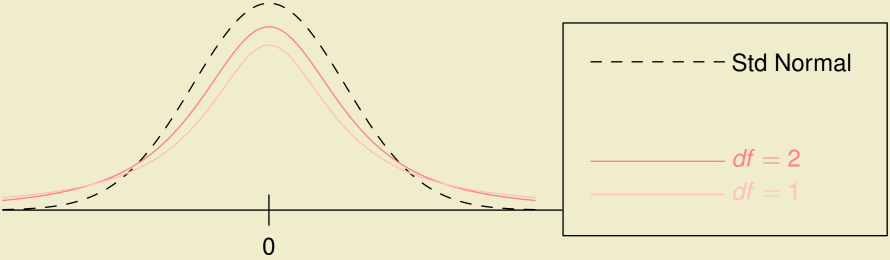
Note

The Student *t* distribution changes for different degrees of freedom.



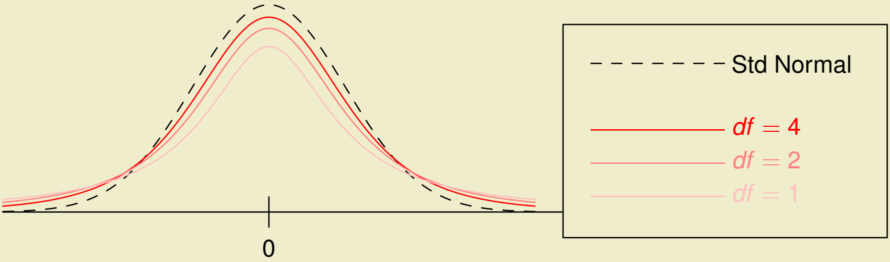
Note

The Student t distribution changes for different degrees of freedom.



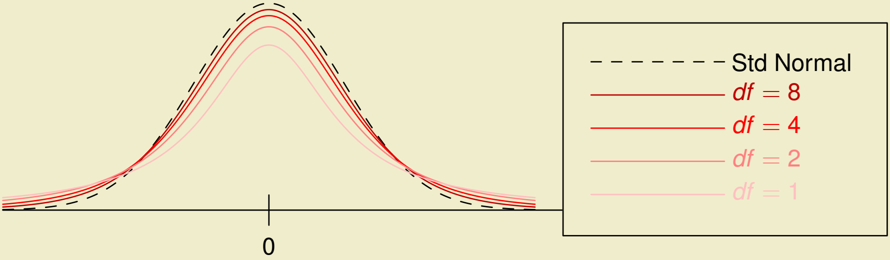
Note

The Student t distribution changes for different degrees of freedom.



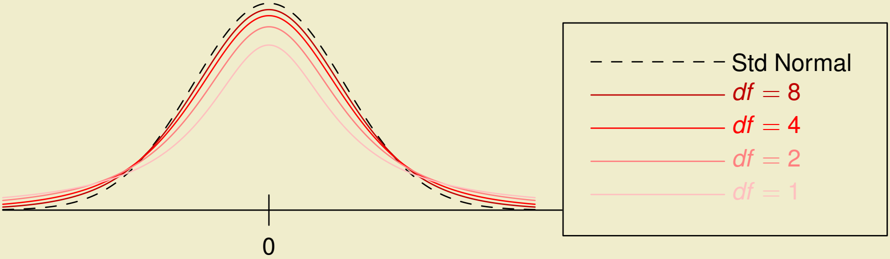
Note

The Student t distribution changes for different degrees of freedom.



Note

The Student t distribution changes for different degrees of freedom.



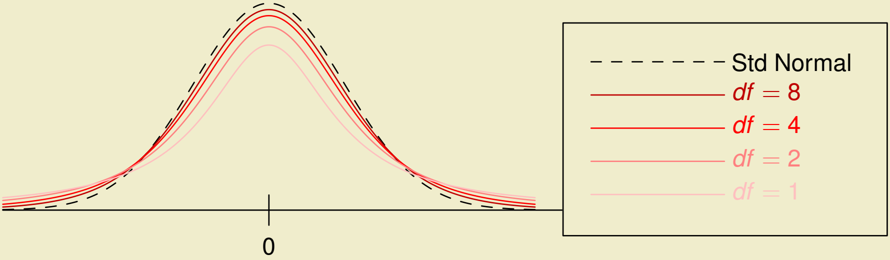
Note

The t-distribution has a mean of $t = 0$

The standard deviation varies with n, but is always greater than 1.

Note

The Student t distribution changes for different degrees of freedom.



Note

The t-distribution has a mean of $t = 0$

The standard deviation varies with n, but is always greater than 1.

Note

As the sample size gets larger, the Student t distribution gets closer to the standard normal distribution.