# AUA CS 108, Statistics, Fall 2019

Lecture 32

Michael Poghosyan
YSU, AUA
michael@ysu.am, mpoghosyan@aua.am

6 Nov 2019

# Contents

# Last Lecture ReCap

- Give the definition of the $(1-\alpha)$-level CI.

# Last Lecture ReCap

- Give the definition of the $(1-\alpha)$-level CI.
- Give the Chebyshev Inequality.

# Last Lecture ReCap

- Give the definition of the $(1-\alpha)$-level CI.
- Give the Chebyshev Inequality.
- What is the **R** command to generate 20 random numbers from the *Cauchy*(2) distribution?

# Last Lecture ReCap

- Give the definition of the $(1-\alpha)$-level CI.
- Give the Chebyshev Inequality.
- What is the **R** command to generate 20 random numbers from the *Cauchy*(2) distribution?
- Give a $(1-\alpha)$-level CI for *p* in *Bernoulli*(*p*) Model.

# CI for the Proportion, Cheby Method

Recall that if we have a Random Sample

$$X_1, X_2, ..., X_n \sim Bernoulli(p),$$

then the interval

$$\left(\overline{X} - \frac{1}{2\sqrt{n \cdot \alpha}},\ \overline{X} + \frac{1}{2\sqrt{n \cdot \alpha}}\right)$$

is a CI for $p$ of level $1 - \alpha$.

# CI for the Proportion, Cheby Method

Recall that if we have a Random Sample

$$X_1, X_2, ..., X_n \sim Bernoulli(p),$$

then the interval

$$\left(\overline{X} - \frac{1}{2\sqrt{n \cdot \alpha}},\ \overline{X} + \frac{1}{2\sqrt{n \cdot \alpha}}\right)$$

is a CI for $p$ of level $1 - \alpha$.

**Note:** Here

$$\frac{1}{2\sqrt{n \cdot \alpha}}$$

is called the **Margin of Error** (for the Interval Estimate of $p$).

# Examples

**Example:** Assume we are interested in the proportion of smokers in AUA. We ask 120 persons at AUA and learn that 55 of them are smokers. Construct a CI for the proportion of smokers in AUA of 95% confidence level.

**Solution:** OTB

# Examples

**Example:** Assume we are interested in the proportion of smokers in AUA. We ask 120 persons at AUA and learn that 55 of them are smokers. Construct a CI for the proportion of smokers in AUA of 95% confidence level.

**Solution:** OTB

**Example:** Continuing the above Example: now assume we want to find that Proportion within the Error Margin 0.1, with the CL 95%. At least, how many persons at AUA we need to ask?

**Solution:** OTB

# CI by Pivotal Quantity Method

Again, we want to construct a CI of CL $1-\alpha$ for $\theta$, using the Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta.$$

# CI by Pivotal Quantity Method

Again, we want to construct a CI of CL $1-\alpha$ for $\theta$, using the Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta.$$

Suppose we can construct a function of $X_1, X_2, ..., X_n$ and our unknown parameter $\theta$,

$$g(X_1, X_2, ..., X_n, \theta)$$

such that

# CI by Pivotal Quantity Method

Again, we want to construct a CI of CL $1-\alpha$ for $\theta$, using the Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta.$$

Suppose we can construct a function of $X_1, X_2, ..., X_n$ and our unknown parameter $\theta$,

$$g(X_1, X_2, ..., X_n, \theta)$$

such that

- the distribution of $g(X_1, ..., X_n, \theta)$ is independent of $\theta$;

# CI by Pivotal Quantity Method

Again, we want to construct a CI of CL $1-\alpha$ for $\theta$, using the Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta.$$

Suppose we can construct a function of $X_1, X_2, ..., X_n$ and our unknown parameter $\theta$,

$$g(X_1, X_2, ..., X_n, \theta)$$

such that

- the distribution of $g(X_1, ..., X_n, \theta)$ is independent of $\theta$;
- $g$ is invertible in terms of $\theta$, say, is strictly monotonic in $\theta$

# CI by Pivotal Quantity Method

Again, we want to construct a CI of CL $1-\alpha$ for $\theta$, using the Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{F}_\theta.$$

Suppose we can construct a function of $X_1, X_2, ..., X_n$ and our unknown parameter $\theta$,

$$g(X_1, X_2, ..., X_n, \theta)$$

such that

- the distribution of $g(X_1, ..., X_n, \theta)$ is independent of $\theta$;
- $g$ is invertible in terms of $\theta$, say, is strictly monotonic in $\theta$

Then we call $g(X_1, ..., X_n, \theta)$ to be a **Pivot** for our model.

# CI by Pivotal Quantity Method, Cont'd

How to use Pivots to construct CIs? The general idea is:

# CI by Pivotal Quantity Method, Cont'd

How to use Pivots to construct CIs? The general idea is:

- Find $a$ and $b$ such that

$$\mathbb{P}\Big(a < g(X_1, X_2, ..., X_n, \theta) < b\Big) \geq 1 - \alpha;$$

## CI by Pivotal Quantity Method, Cont'd

How to use Pivots to construct CIs? The general idea is:

- Find $a$ and $b$ such that

$$\mathbb{P}\Big(a < g(X_1, X_2, ..., X_n, \theta) < b\Big) \geq 1 - \alpha;$$

  Because of the first condition, $a$ and $b$ will not depend on $\theta$;

# CI by Pivotal Quantity Method, Cont'd

How to use Pivots to construct CIs? The general idea is:

- Find $a$ and $b$ such that
$$\mathbb{P}\Big(a < g(X_1, X_2, ..., X_n, \theta) < b\Big) \geq 1 - \alpha;$$
Because of the first condition, $a$ and $b$ will not depend on $\theta$;
- now solve the inequality
$$a < g(X_1, ..., X_n, \theta) < b$$
for $\theta$ for the obtained $a$ and $b$.

# CI by Pivotal Quantity Method, Cont'd

How to use Pivots to construct CIs? The general idea is:

- Find $a$ and $b$ such that
$$\mathbb{P}\Big(a < g(X_1, X_2, ..., X_n, \theta) < b\Big) \geq 1 - \alpha;$$
Because of the first condition, $a$ and $b$ will not depend on $\theta$;
- now solve the inequality
$$a < g(X_1, ..., X_n, \theta) < b$$
for $\theta$ for the obtained $a$ and $b$.

This will give some inequality of the form $L < \theta < U$, and hence,
$$\mathbb{P}(L < \theta < U) \geq 1 - \alpha,$$
so we can take $(L, U)$ as a CI.

# CI by Pivotal Quantity Method, Cont'd

How to use Pivots to construct CIs? The general idea is:

- Find $a$ and $b$ such that
$$\mathbb{P}\Big(a < g(X_1, X_2, ..., X_n, \theta) < b\Big) \geq 1 - \alpha;$$
Because of the first condition, $a$ and $b$ will not depend on $\theta$;
- now solve the inequality
$$a < g(X_1, ..., X_n, \theta) < b$$
for $\theta$ for the obtained $a$ and $b$.

This will give some inequality of the form $L < \theta < U$, and hence,
$$\mathbb{P}(L < \theta < U) \geq 1 - \alpha,$$
so we can take $(L, U)$ as a CI.

**Note:** Usually, we are solving
$$\mathbb{P}\Big(a < g(X_1, X_2, ..., X_n, \theta) < b\Big) = 1 - \alpha.$$

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

**Problem:** Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Assume $\sigma^2$ is known. Given $\alpha \in (0, 1)$, we want to construct a CI of CL $1 - \alpha$ for $\mu$, using a Pivot.

## CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

**Problem:** Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Assume $\sigma^2$ is known. Given $\alpha \in (0, 1)$, we want to construct a CI of CL $1 - \alpha$ for $\mu$, using a Pivot.

**Solution:** We know that $\overline{X}$ is a good Estimator for $\mu$.

## CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

**Problem:** Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Assume $\sigma^2$ is known. Given $\alpha \in (0, 1)$, we want to construct a CI of CL $1 - \alpha$ for $\mu$, using a Pivot.

**Solution:** We know that $\overline{X}$ is a good Estimator for $\mu$. So we will use it to obtain a Pivot.

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

**Problem:** Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Assume $\sigma^2$ is known. Given $\alpha \in (0, 1)$, we want to construct a CI of CL $1 - \alpha$ for $\mu$, using a Pivot.

**Solution:** We know that $\overline{X}$ is a good Estimator for $\mu$. So we will use it to obtain a Pivot.

We know that, since $X_k \sim \mathcal{N}(\mu, \sigma^2)$ are IID, then

$$\overline{X} \sim$$

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

**Problem:** Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Assume $\sigma^2$ is known. Given $\alpha \in (0, 1)$, we want to construct a CI of CL $1 - \alpha$ for $\mu$, using a Pivot.

**Solution:** We know that $\overline{X}$ is a good Estimator for $\mu$. So we will use it to obtain a Pivot.

We know that, since $X_k \sim \mathcal{N}(\mu, \sigma^2)$ are IID, then

$$\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

## CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

**Problem:** Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Assume $\sigma^2$ is known. Given $\alpha \in (0, 1)$, we want to construct a CI of CL $1 - \alpha$ for $\mu$, using a Pivot.

**Solution:** We know that $\overline{X}$ is a good Estimator for $\mu$. So we will use it to obtain a Pivot.

We know that, since $X_k \sim \mathcal{N}(\mu, \sigma^2)$ are IID, then

$$\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

This shows that $\overline{X}$ cannot serve as a Pivot, since its Distribution depends on the unknown $\mu$.

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

**Problem:** Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Assume $\sigma^2$ is known. Given $\alpha \in (0, 1)$, we want to construct a CI of CL $1 - \alpha$ for $\mu$, using a Pivot.

**Solution:** We know that $\overline{X}$ is a good Estimator for $\mu$. So we will use it to obtain a Pivot.

We know that, since $X_k \sim \mathcal{N}(\mu, \sigma^2)$ are IID, then

$$\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

This shows that $\overline{X}$ cannot serve as a Pivot, since its Distribution depends on the unknown $\mu$.

But let us consider $\overline{X} - \mu$.

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

Clearly,

$$\overline{X} - \mu \sim$$

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

Clearly,

$$\overline{X} - \mu \sim \mathcal{N}\left(0, \frac{\sigma^2}{n}\right).$$

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

Clearly,

$$\overline{X} - \mu \sim \mathcal{N}\left(0, \frac{\sigma^2}{n}\right).$$

So the Distribution of $\overline{X} - \mu$ is *independent of* $\mu$!

## CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

Clearly,

$$\overline{X} - \mu \sim \mathcal{N}\left(0, \frac{\sigma^2}{n}\right).$$

So the Distribution of $\overline{X} - \mu$ is *independent of* $\mu$! Uraaa! We found a Pivoooot.

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

Clearly,

$$\overline{X} - \mu \sim \mathcal{N}\left(0, \frac{\sigma^2}{n}\right).$$

So the Distribution of $\overline{X} - \mu$ is *independent of* $\mu$! Uraaa! We found a Pivoooot. In fact, usually people take the following Pivot:

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim$$

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

Clearly,

$$\overline{X} - \mu \sim \mathcal{N}\left(0, \frac{\sigma^2}{n}\right).$$

So the Distribution of $\overline{X} - \mu$ is *independent of* $\mu$! Uraaa! We found a Pivoooot. In fact, usually people take the following Pivot:

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

since working with $\mathcal{N}(0, 1)$ is easier: we have Standard Tables for Standard Normal!

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

Clearly,

$$\overline{X} - \mu \sim \mathcal{N}\left(0, \frac{\sigma^2}{n}\right).$$

So the Distribution of $\overline{X} - \mu$ is *independent of* $\mu$! Uraaa! We found a Pivoooot. In fact, usually people take the following Pivot:

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

since working with $\mathcal{N}(0, 1)$ is easier: we have Standard Tables for Standard Normal! So what??? Who cares for that in this digital era of the 4th Industrial Revolution? $\ddot{\smile}$

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

Clearly,

$$\overline{X} - \mu \sim \mathcal{N}\left(0, \frac{\sigma^2}{n}\right).$$

So the Distribution of $\overline{X} - \mu$ is *independent of* $\mu$! Uraaa! We found a Pivoooot. In fact, usually people take the following Pivot:

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

since working with $\mathcal{N}(0, 1)$ is easier: we have Standard Tables for Standard Normal! So what??? Who cares for that in this digital era of the 4th Industrial Revolution? $\ddot\smile$

The name of the above ratio is $Z$-**statistics**, and we will meet this again in Hypotheses testing part.

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

Now, let us use this Pivot: we have $Z \sim \mathcal{N}(0, 1)$.

## CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

Now, let us use this Pivot: we have $Z \sim \mathcal{N}(0, 1)$. Let us find numbers $a$ and $b$ such that

$$\mathbb{P}(a < Z < b) = 1 - \alpha.$$

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

Now, let us use this Pivot: we have $Z \sim \mathcal{N}(0,1)$. Let us find numbers $a$ and $b$ such that

$$\mathbb{P}(a < Z < b) = 1 - \alpha.$$

This means that we want to select an interval, over which the area under the Standard Normal PDF is $1 - \alpha$.

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

Now, let us use this Pivot: we have $Z \sim \mathcal{N}(0,1)$. Let us find numbers $a$ and $b$ such that

$$\mathbb{P}(a < Z < b) = 1 - \alpha.$$

This means that we want to select an interval, over which the area under the Standard Normal PDF is $1 - \alpha$. Clearly, there exist infinitely many this kind of intervals.

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

Now, let us use this Pivot: we have $Z \sim \mathcal{N}(0, 1)$. Let us find numbers $a$ and $b$ such that

$$\mathbb{P}(a < Z < b) = 1 - \alpha.$$

This means that we want to select an interval, over which the area under the Standard Normal PDF is $1 - \alpha$. Clearly, there exist infinitely many this kind of intervals. But let us choose the *minimum length* interval, which will be

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

Now, let us use this Pivot: we have $Z \sim \mathcal{N}(0,1)$. Let us find numbers $a$ and $b$ such that

$$\mathbb{P}(a < Z < b) = 1 - \alpha.$$

This means that we want to select an interval, over which the area under the Standard Normal PDF is $1 - \alpha$. Clearly, there exist infinitely many this kind of intervals. But let us choose the *minimum length* interval, which will be a symmetric interval around 0 (prove!).

## CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

Now, let us use this Pivot: we have $Z \sim \mathcal{N}(0,1)$. Let us find numbers $a$ and $b$ such that

$$\mathbb{P}(a < Z < b) = 1 - \alpha.$$

This means that we want to select an interval, over which the area under the Standard Normal PDF is $1 - \alpha$. Clearly, there exist infinitely many this kind of intervals. But let us choose the *minimum length* interval, which will be a symmetric interval around 0 (prove!). So we need to find an inteval of the form $(-b, b)$ such that

$$\mathbb{P}(-b < Z < b) = 1 - \alpha.$$

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

Now, let us use this Pivot: we have $Z \sim \mathcal{N}(0, 1)$. Let us find numbers $a$ and $b$ such that

$$\mathbb{P}(a < Z < b) = 1 - \alpha.$$

This means that we want to select an interval, over which the area under the Standard Normal PDF is $1 - \alpha$. Clearly, there exist infinitely many this kind of intervals. But let us choose the *minimum length* interval, which will be a symmetric interval around 0 (prove!). So we need to find an inteval of the form $(-b, b)$ such that

$$\mathbb{P}(-b < Z < b) = 1 - \alpha.$$

Because of the symmetry, we need to have that the area of both tails is $\alpha/2$:

$$\mathbb{P}(Z \leq -b) = \mathbb{P}(Z \geq b) = \frac{\alpha}{2},$$

hence,

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

Now, let us use this Pivot: we have $Z \sim \mathcal{N}(0, 1)$. Let us find numbers $a$ and $b$ such that

$$\mathbb{P}(a < Z < b) = 1 - \alpha.$$

This means that we want to select an interval, over which the area under the Standard Normal PDF is $1 - \alpha$. Clearly, there exist infinitely many this kind of intervals. But let us choose the *minimum length* interval, which will be a symmetric interval around 0 (prove!). So we need to find an inteval of the form $(-b, b)$ such that

$$\mathbb{P}(-b < Z < b) = 1 - \alpha.$$

Because of the symmetry, we need to have that the area of both tails is $\alpha/2$:

$$\mathbb{P}(Z \leq -b) = \mathbb{P}(Z \geq b) = \frac{\alpha}{2},$$

hence,

$$b = z_{1-\frac{\alpha}{2}},$$

where $z_{1-\alpha/2}$ is the $1 - \frac{\alpha}{2}$ quantile of the Standard Normal Distribution.

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

So we obtained

$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha.$$

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

So we obtained

$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha.$$

We plug here the value of $Z$:

$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha,$$

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

So we obtained

$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha.$$

We plug here the value of $Z$:

$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha,$$

and solve for $\mu$:

$$\mathbb{P}\left(\overline{X} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is known, Pivotal Method

So we obtained

$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha.$$

We plug here the value of $Z$:

$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha,$$

and solve for $\mu$:

$$\mathbb{P}\left(\overline{X} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

Hence, the following interval:

$$\left(\overline{X} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}},\ \ \overline{X} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

is a $(1-\alpha)$-level CI for $\mu$.

# Example

**Example:** Assume we want to construct a 95% CI for $\mu$ in the $\mathcal{N}(\mu, \sigma^2)$ Model, when $\sigma$ is given, known.

# Example

**Example:** Assume we want to construct a 95% CI for $\mu$ in the $\mathcal{N}(\mu, \sigma^2)$ Model, when $\sigma$ is given, known. We construct it using a Random Sample

$$X_1, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

# Example

**Example:** Assume we want to construct a 95% CI for $\mu$ in the $\mathcal{N}(\mu, \sigma^2)$ Model, when $\sigma$ is given, known. We construct it using a Random Sample

$$X_1, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Since $1 - \alpha = 0.95$, then $\alpha = 0.05$.

# Example

**Example:** Assume we want to construct a 95% CI for $\mu$ in the $\mathcal{N}(\mu, \sigma^2)$ Model, when $\sigma$ is given, known. We construct it using a Random Sample

$$X_1, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Since $1 - \alpha = 0.95$, then $\alpha = 0.05$. By the above formula, we need to calculate the Standard Normal quantile $z_{1-\alpha/2} =$

# Example

**Example:** Assume we want to construct a 95% CI for $\mu$ in the $\mathcal{N}(\mu, \sigma^2)$ Model, when $\sigma$ is given, known. We construct it using a Random Sample

$$X_1, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Since $1 - \alpha = 0.95$, then $\alpha = 0.05$. By the above formula, we need to calculate the Standard Normal quantile $z_{1-\alpha/2} = z_{0.975}$.

# Example

**Example:** Assume we want to construct a 95% CI for $\mu$ in the $\mathcal{N}(\mu, \sigma^2)$ Model, when $\sigma$ is given, known. We construct it using a Random Sample

$$X_1, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Since $1 - \alpha = 0.95$, then $\alpha = 0.05$. By the above formula, we need to calculate the Standard Normal quantile $z_{1-\alpha/2} = z_{0.975}$.

**R** gives:

```r
qnorm(0.975)
```

```
## [1] 1.959964
```

so our 95% CI will be

$$\left(\overline{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \;\; \overline{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right).$$

# Example

**Example with R:** We generate random numbers from $\mathcal{N}(2.31, 4)$ (so here we assume we know the true parameter value of $\mu$).

```
sigma <- 2
n <- 20
smpl <- rnorm(n, mean = 2.31, sd = sigma)
smpl
```

```
##  [1]  2.4067314  6.7984510  3.5772345  1.5963469 -1.894
##  [7]  3.0240236  2.9426280  3.7549903  4.4298303  3.125
## [13]  0.7173930  4.9879032  2.8020583  2.9063363  1.931
## [19]  3.4198754  1.0501713
```

# Example

**Example with R:** We generate random numbers from $\mathcal{N}(2.31, 4)$ (so here we assume we know the true parameter value of $\mu$).

```
sigma <- 2
n <- 20
smpl <- rnorm(n, mean = 2.31, sd = sigma)
smpl
```

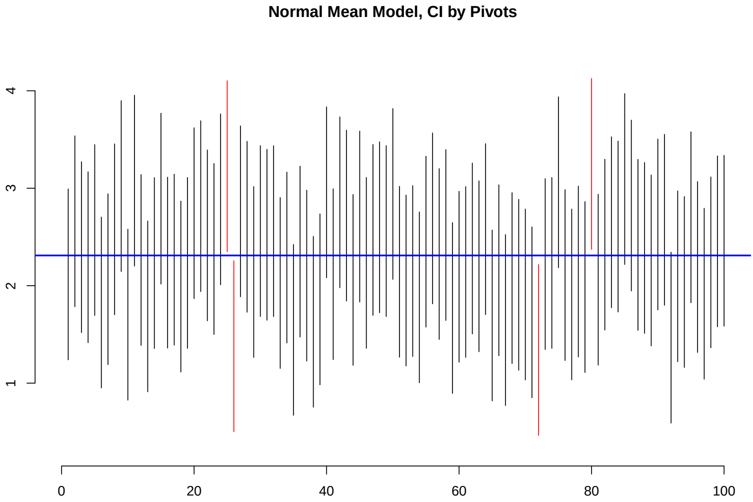
```
##  [1]  2.4067314  6.7984510  3.5772345  1.5963469 -1.894
##  [7]  3.0240236  2.9426280  3.7549903  4.4298303  3.125
## [13]  0.7173930  4.9879032  2.8020583  2.9063363  1.931
## [19]  3.4198754  1.0501713
```

Now we construct the 95% CI by the above formula:

```
me <- 1.96* sigma/sqrt(n) #the Margin of Error
c(mean(smpl) - me, mean(smpl) + me)
```

```
## [1] 1.764134 3.517211
```

# Example, Simulation



# Example, Simulation, Code

```
mu <- 2.31; sigma <- 2
conf.level <- 0.95; a = 1 - conf.level
sample.size <- 20; no.of.intervals <- 100
z <- qnorm(1-a/2) ## our quantile
ME <- z*sigma/sqrt(sample.size) #Margin of Error

plot.new()
plot.window(xlim = c(0,no.of.intervals), ylim = c(mu-2,mu+2))
axis(1); axis(2)
title("Normal Mean Model, CI by Pivots")
for(i in 1:no.of.intervals){
  x <- rnorm(sample.size, mean = mu, sd = sigma)
  lo <- mean(x) - ME; up <- mean(x) + ME
  if(lo > mu || up < mu){
    segments(c(i), c(lo), c(i), c(up), col = "red")
  }
  else{
    segments(c(i), c(lo), c(i), c(up))
  }
}
abline(h = mu, lwd = 2, col = "blue")
```

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is **unknown**, PivMe

**Problem:** Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Assume $\sigma^2$ is **unknown**, which is more realistic.

## CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is **unknown**, PivMe

**Problem:** Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Assume $\sigma^2$ is **unknown**, which is more realistic. Given $\alpha \in (0, 1)$, we want to construct a CI of CL $1 - \alpha$ for $\mu$, using a Pivot.

## CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is **unknown**, PivMe

**Problem:** Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Assume $\sigma^2$ is **unknown**, which is more realistic. Given $\alpha \in (0, 1)$, we want to construct a CI of CL $1 - \alpha$ for $\mu$, using a Pivot.

**Solution:** Again we start from $\overline{X}$, having that it is a good Estimator for $\mu$. Again, from the fact that $X_k \sim \mathcal{N}(\mu, \sigma^2)$ are IID, we will have

$$\overline{X} \sim$$

## CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is **unknown**, PivMe

**Problem:** Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Assume $\sigma^2$ is **unknown**, which is more realistic. Given $\alpha \in (0, 1)$, we want to construct a CI of CL $1 - \alpha$ for $\mu$, using a Pivot.

**Solution:** Again we start from $\overline{X}$, having that it is a good Estimator for $\mu$. Again, from the fact that $X_k \sim \mathcal{N}(\mu, \sigma^2)$ are IID, we will have

$$\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is **unknown**, PivMe

**Problem:** Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Assume $\sigma^2$ is **unknown**, which is more realistic. Given $\alpha \in (0, 1)$, we want to construct a CI of CL $1 - \alpha$ for $\mu$, using a Pivot.

**Solution:** Again we start from $\overline{X}$, having that it is a good Estimator for $\mu$. Again, from the fact that $X_k \sim \mathcal{N}(\mu, \sigma^2)$ are IID, we will have

$$\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

In this case, again

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim$$

## CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is **unknown**, PivMe

**Problem:** Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Assume $\sigma^2$ is **unknown**, which is more realistic. Given $\alpha \in (0, 1)$, we want to construct a CI of CL $1 - \alpha$ for $\mu$, using a Pivot.

**Solution:** Again we start from $\overline{X}$, having that it is a good Estimator for $\mu$. Again, from the fact that $X_k \sim \mathcal{N}(\mu, \sigma^2)$ are IID, we will have

$$\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

In this case, again

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1),$$

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is **unknown**, PivMe

**Problem:** Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Assume $\sigma^2$ is **unknown**, which is more realistic. Given $\alpha \in (0, 1)$, we want to construct a CI of CL $1 - \alpha$ for $\mu$, using a Pivot.

**Solution:** Again we start from $\overline{X}$, having that it is a good Estimator for $\mu$. Again, from the fact that $X_k \sim \mathcal{N}(\mu, \sigma^2)$ are IID, we will have

$$\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

In this case, again

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1),$$

but, unfortunately, we cannot use $Z$, since the result will contain $\sigma$, **which is unknown to us**.

## CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is **unknown**, PivMe

**Problem:** Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Assume $\sigma^2$ is **unknown**, which is more realistic. Given $\alpha \in (0, 1)$, we want to construct a CI of CL $1 - \alpha$ for $\mu$, using a Pivot.

**Solution:** Again we start from $\overline{X}$, having that it is a good Estimator for $\mu$. Again, from the fact that $X_k \sim \mathcal{N}(\mu, \sigma^2)$ are IID, we will have

$$\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

In this case, again

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1),$$

but, unfortunately, we cannot use $Z$, since the result will contain $\sigma$, **which is unknown to us**. So we need to adjust $Z$.

## CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is **unknown**, PivMe

We know some good Estimators for $\sigma$: let us take, in this case, the following version of Sample SD:

$$S^2 = \frac{\sum_{k=1}^{n}(X_k - \overline{X})^2}{n-1}, \qquad i.e., \qquad S = \sqrt{\frac{\sum_{k=1}^{n}(X_k - \overline{X})^2}{n-1}}.$$

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is **unknown**, PivMe

We know some good Estimators for $\sigma$: let us take, in this case, the following version of Sample SD:

$$S^2 = \frac{\sum_{k=1}^{n}(X_k - \overline{X})^2}{n-1}, \qquad i.e., \qquad S = \sqrt{\frac{\sum_{k=1}^{n}(X_k - \overline{X})^2}{n-1}}.$$

Now, let us replace above $\sigma$ with $S$:

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}}.$$

## CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is **unknown**, PivMe

We know some good Estimators for $\sigma$: let us take, in this case, the following version of Sample SD:

$$S^2 = \frac{\sum_{k=1}^n (X_k - \overline{X})^2}{n-1}, \qquad i.e., \qquad S = \sqrt{\frac{\sum_{k=1}^n (X_k - \overline{X})^2}{n-1}}.$$

Now, let us replace above $\sigma$ with $S$:

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}}.$$

OK, what's next? Is it a Pivot? Is the Distribution of $t$ independent of $\mu$?

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is **unknown**, PivMe

We know some good Estimators for $\sigma$: let us take, in this case, the following version of Sample SD:

$$S^2 = \frac{\sum_{k=1}^n (X_k - \overline{X})^2}{n-1}, \qquad i.e., \qquad S = \sqrt{\frac{\sum_{k=1}^n (X_k - \overline{X})^2}{n-1}}.$$

Now, let us replace above $\sigma$ with $S$:

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}}.$$

OK, what's next? Is it a Pivot? Is the Distribution of $t$ independent of $\mu$? Or, in general, do we know the Distribution of $t$ ?

I know, not sure about you.

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is **unknown**, PivMe

We know some good Estimators for $\sigma$: let us take, in this case, the following version of Sample SD:

$$S^2 = \frac{\sum_{k=1}^{n}(X_k - \overline{X})^2}{n-1}, \qquad i.e., \qquad S = \sqrt{\frac{\sum_{k=1}^{n}(X_k - \overline{X})^2}{n-1}}.$$

Now, let us replace above $\sigma$ with $S$:

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}}.$$

OK, what's next? Is it a Pivot? Is the Distribution of $t$ independent of $\mu$? Or, in general, do we know the Distribution of $t$ ?

I know, not sure about you. Yet!

# $t$-Distribution

It turns out that the Distribution of above

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

is the famous $t$-Distribution with $n - 1$ degrees of freedom:

[1]See, e.g., https://en.wikipedia.org/wiki/Student's_t-distribution

# $t$-Distribution

It turns out that the Distribution of above

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

is the famous $t$-Distribution with $n - 1$ degrees of freedom:

**Definition:** If $X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2)$ are IID and

$$S^2 = \frac{\sum_{k=1}^{n}(X_k - \overline{X})^2}{n-1},$$

then the Distribution of

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

is called the **Student's $t$-Distribution with $n - 1$ degrees of freedom**[1], and is denoted by $t(n-1)$.

[1] See, e.g., https://en.wikipedia.org/wiki/Student's_t-distribution

Have you ever heard the term *Vechniy Student* ?

Have you ever heard the term *Vechniy Student* ? Here is he:

Have you ever heard the term *Vechniy Student* ? Here is he:

Have you ever heard the term *Vechniy Student* ? Here is he:

British statistician William Sealy Gosset, known as "Student"

Have you ever heard the term *Vechniy Student* ? Here is he:

British statistician William Sealy Gosset, known as "Student"

You can read about him at
https://en.wikipedia.org/wiki/William_Sealy_Gosset.

# Student's Paper

# Student's Paper

VOLUME VI MARCH, 1908 No. 1

BIOMETRIKA.

## THE PROBABLE ERROR OF A MEAN.

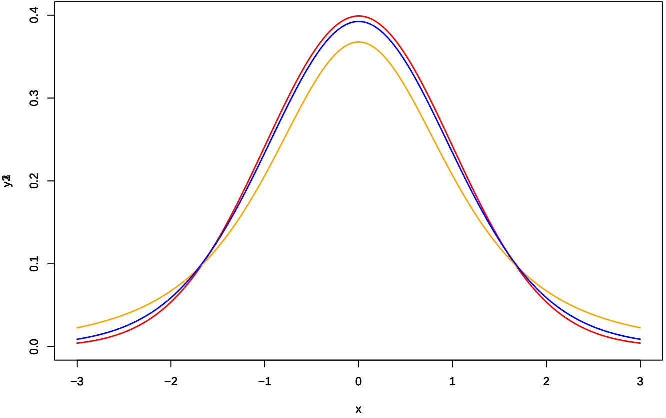
BY STUDENT.

*Introduction.*

ANY experiment may be regarded as forming an individual of a "population" of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population.

# t-Distribution

```
x <- seq(-3,3, 0.01)
y1 <- dnorm(x); y2 <- dt(x, df = 3); y3 <- dt(x, df = 15)
plot(x,y1, type = "l", col = "red", lwd = 2, ylim = c(0, 0.4))
par(new = T)
plot(x,y2, type = "l", col = "orange", lwd = 2, ylim = c(0, 0.4))
par(new = T)
plot(x,y3, type = "l", col = "blue", lwd = 2, ylim = c(0, 0.4))
```



## CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is **unknown**, PivMe

Back to our Problem, we know that

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1),$$

and the Distribution of $t$ is independent of $\mu$, so it is a Pivot for $\mu$.

## CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is **unknown**, PivMe

Back to our Problem, we know that

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1),$$

and the Distribution of $t$ is independent of $\mu$, so it is a Pivot for $\mu$. We call it a $t$-**statistics**.

## CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is **unknown**, PivMe

Back to our Problem, we know that

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1),$$

and the Distribution of $t$ is independent of $\mu$, so it is a Pivot for $\mu$. We call it a $t$-**statistics**.

Now, for $t \sim t(n-1)$, let us find numbers $a$ and $b$ such that

$$\mathbb{P}(a < t < b) = 1 - \alpha.$$

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is **unknown**, PivMe

Back to our Problem, we know that

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1),$$

and the Distribution of $t$ is independent of $\mu$, so it is a Pivot for $\mu$. We call it a $t$-**statistics**.

Now, for $t \sim t(n-1)$, let us find numbers $a$ and $b$ such that

$$\mathbb{P}(a < t < b) = 1 - \alpha.$$

Again, because of the symmetry of the PDF of $t(n-1)$, the *minimum length* interval will be symmetric around 0.

## CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is **unknown**, PivMe

Back to our Problem, we know that

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1),$$

and the Distribution of $t$ is independent of $\mu$, so it is a Pivot for $\mu$. We call it a $t$-**statistics**.

Now, for $t \sim t(n-1)$, let us find numbers $a$ and $b$ such that

$$\mathbb{P}(a < t < b) = 1 - \alpha.$$

Again, because of the symmetry of the PDF of $t(n-1)$, the *minimum length* interval will be symmetric around 0. So we need to find an inteval of the form $(-b, b)$ such that

$$\mathbb{P}(-b < t < b) = 1 - \alpha.$$

## CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is **unknown**, PivMe

Back to our Problem, we know that

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1),$$

and the Distribution of $t$ is independent of $\mu$, so it is a Pivot for $\mu$. We call it a $t$-**statistics**.

Now, for $t \sim t(n-1)$, let us find numbers $a$ and $b$ such that

$$\mathbb{P}(a < t < b) = 1 - \alpha.$$

Again, because of the symmetry of the PDF of $t(n-1)$, the *minimum length* interval will be symmetric around 0. So we need to find an inteval of the form $(-b, b)$ such that

$$\mathbb{P}(-b < t < b) = 1 - \alpha.$$

Again, because of the symmetry, we need to have that the area of both tails is $\alpha/2$:

$$\mathbb{P}(t \leq -b) = \mathbb{P}(t \geq b) = \frac{\alpha}{2}.$$

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is **unknown**, PivMe

Hence,

## CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is **unknown**, PivMe

Hence,

$$b = t_{1-\frac{\alpha}{2}}(n-1) = t_{n-1,1-\frac{\alpha}{2}},$$

where $t_{n-1,1-\alpha/2}$ is the $1-\frac{\alpha}{2}$ quantile of the $t(n-1)$ Distribution.

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is **unknown**, PivMe

Hence,

$$b = t_{1-\frac{\alpha}{2}}(n-1) = t_{n-1,1-\frac{\alpha}{2}},$$

where $t_{n-1,1-\alpha/2}$ is the $1-\frac{\alpha}{2}$ quantile of the $t(n-1)$ Distribution.

So we obtained

$$\mathbb{P}\left(-t_{n-1,1-\alpha/2} < t < t_{n-1,1-\alpha/2}\right) = 1-\alpha.$$

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is **unknown**, PivMe

Hence,

$$b = t_{1-\frac{\alpha}{2}}(n-1) = t_{n-1,1-\frac{\alpha}{2}},$$

where $t_{n-1,1-\alpha/2}$ is the $1-\frac{\alpha}{2}$ quantile of the $t(n-1)$ Distribution.

So we obtained

$$\mathbb{P}\left(-t_{n-1,1-\alpha/2} < t < t_{n-1,1-\alpha/2}\right) = 1-\alpha.$$

We plug here the value of $t$:

$$\mathbb{P}\left(-t_{n-1,1-\alpha/2} < \frac{\overline{X}-\mu}{S/\sqrt{n}} < t_{n-1,1-\alpha/2}\right) = 1-\alpha,$$

# CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is **unknown**, PivMe

Hence,

$$b = t_{1-\frac{\alpha}{2}}(n-1) = t_{n-1,1-\frac{\alpha}{2}},$$

where $t_{n-1,1-\alpha/2}$ is the $1 - \frac{\alpha}{2}$ quantile of the $t(n-1)$ Distribution.

So we obtained

$$\mathbb{P}\left(-t_{n-1,1-\alpha/2} < t < t_{n-1,1-\alpha/2}\right) = 1 - \alpha.$$

We plug here the value of $t$:

$$\mathbb{P}\left(-t_{n-1,1-\alpha/2} < \frac{\overline{X} - \mu}{S/\sqrt{n}} < t_{n-1,1-\alpha/2}\right) = 1 - \alpha,$$

and solve for $\mu$:

$$\mathbb{P}\left(\overline{X} - t_{n-1,1-\alpha/2} \cdot \frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{n-1,1-\alpha/2} \cdot \frac{S}{\sqrt{n}}\right) = 1 - \alpha.$$

## CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, $\sigma$ is **unknown**, PivMe

Hence,

$$b = t_{1-\frac{\alpha}{2}}(n-1) = t_{n-1,1-\frac{\alpha}{2}},$$

where $t_{n-1,1-\alpha/2}$ is the $1-\frac{\alpha}{2}$ quantile of the $t(n-1)$ Distribution.

So we obtained

$$\mathbb{P}\left(-t_{n-1,1-\alpha/2} < t < t_{n-1,1-\alpha/2}\right) = 1-\alpha.$$

We plug here the value of $t$:

$$\mathbb{P}\left(-t_{n-1,1-\alpha/2} < \frac{\overline{X}-\mu}{S/\sqrt{n}} < t_{n-1,1-\alpha/2}\right) = 1-\alpha,$$

and solve for $\mu$:

$$\mathbb{P}\left(\overline{X} - t_{n-1,1-\alpha/2}\cdot\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{n-1,1-\alpha/2}\cdot\frac{S}{\sqrt{n}}\right) = 1-\alpha.$$

Hence, the following interval:

$$\left(\overline{X} - t_{n-1,1-\alpha/2}\cdot\frac{S}{\sqrt{n}},\ \ \overline{X} + t_{n-1,1-\alpha/2}\cdot\frac{S}{\sqrt{n}}\right)$$

is a $(1-\alpha)$-level CI for $\mu$.