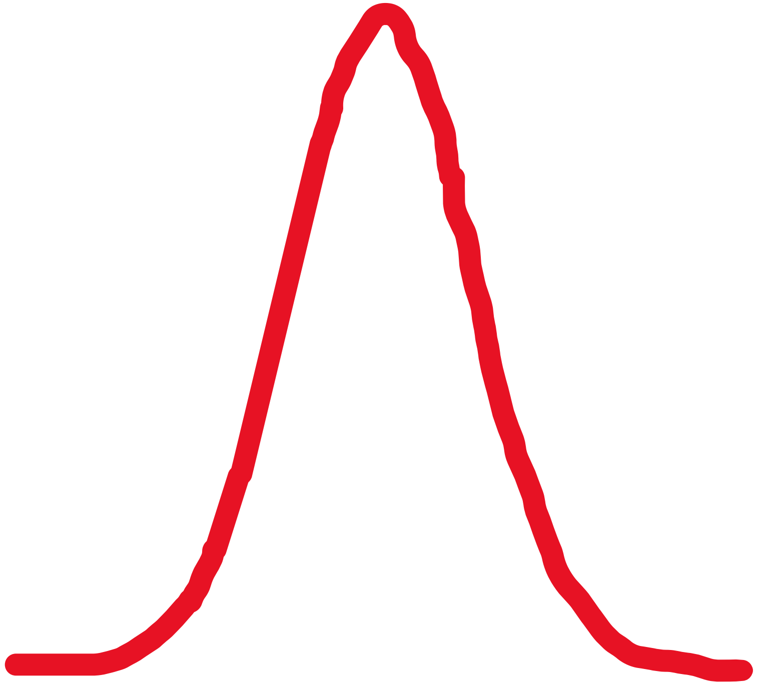
17 – Confidence Intervals (unknown σ)

Reference: [ES] 6.2

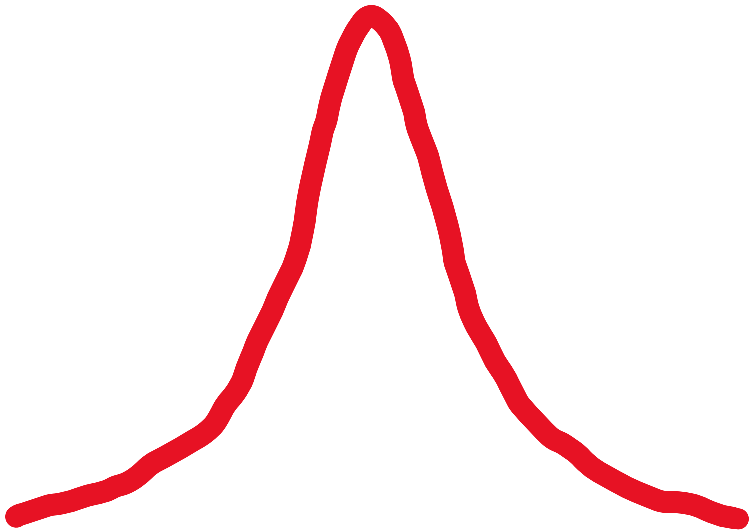
Sampling Distribution

Given pop. distribution with mean μ,
the sampling distribution of size n looks like:

For σ known:



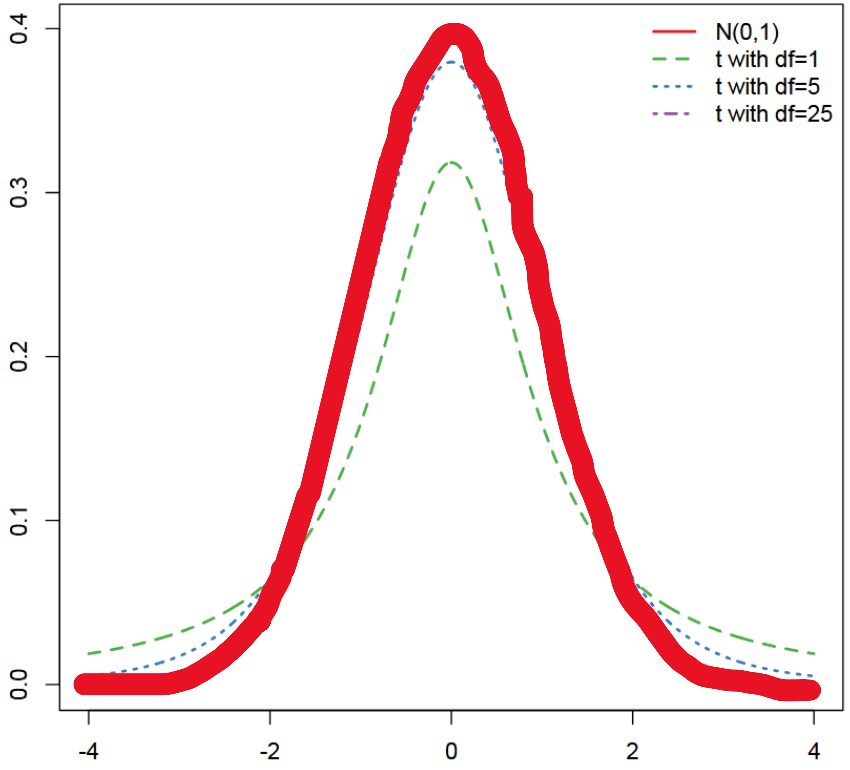
For σ unknown:
Use sample mean s_x to approximate σ.



Sampling Distribution

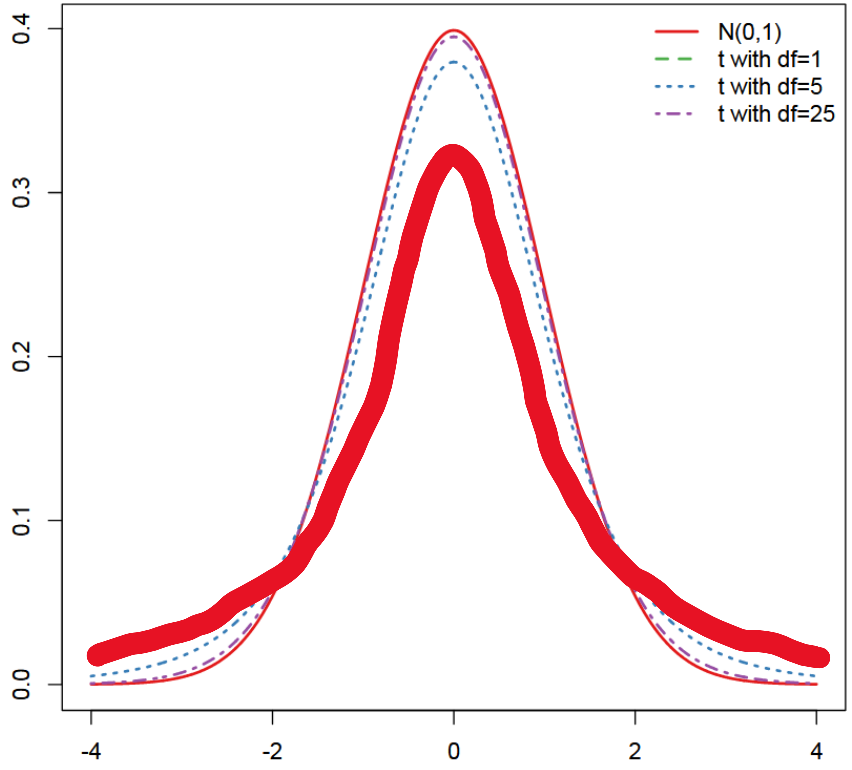
Given pop. distribution with mean μ,
the sampling distribution of size n looks like:

For σ known:



Sampling distribution of $\bar{x}$ is normal with $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

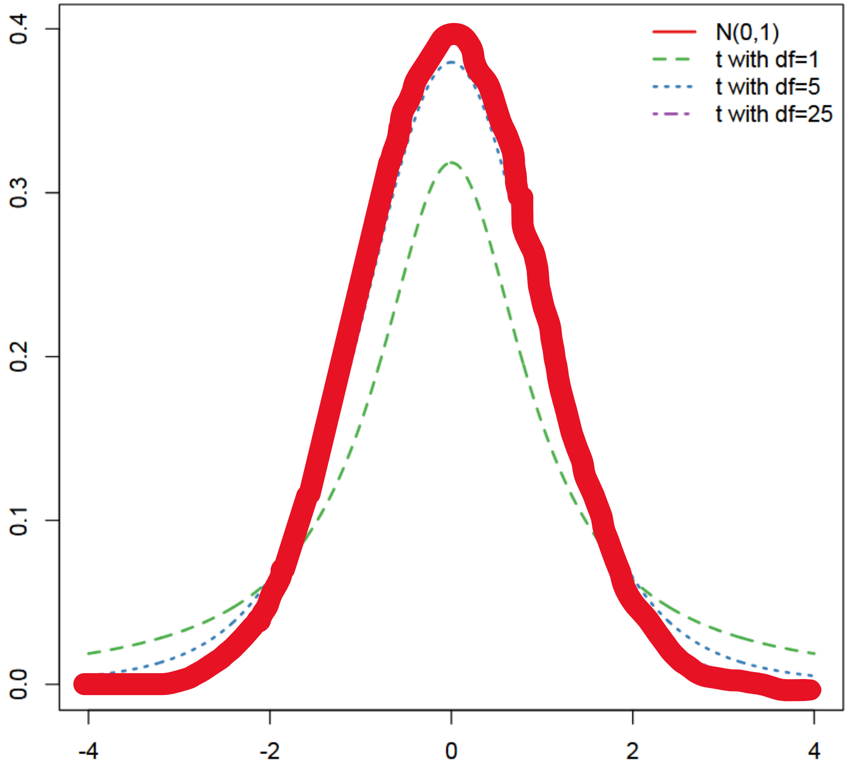
For σ unknown:
Use sample mean s_x to approximate σ.



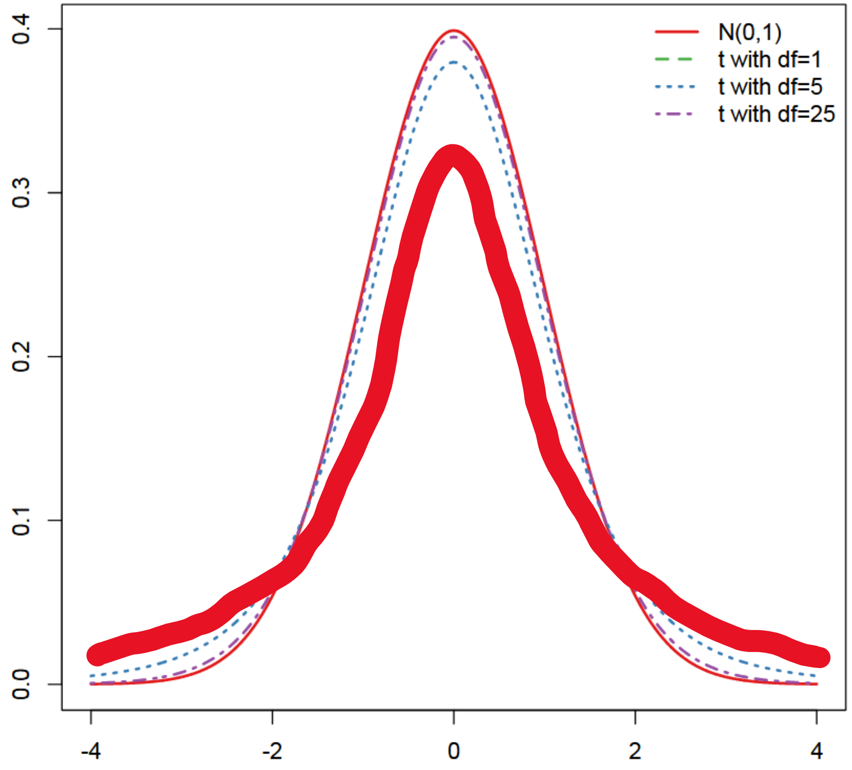
Sampling Distribution

Given pop. distribution with mean μ,
the sampling distribution of size n looks like:

For σ known:



For σ unknown:
Use sample mean s_x to approximate σ.

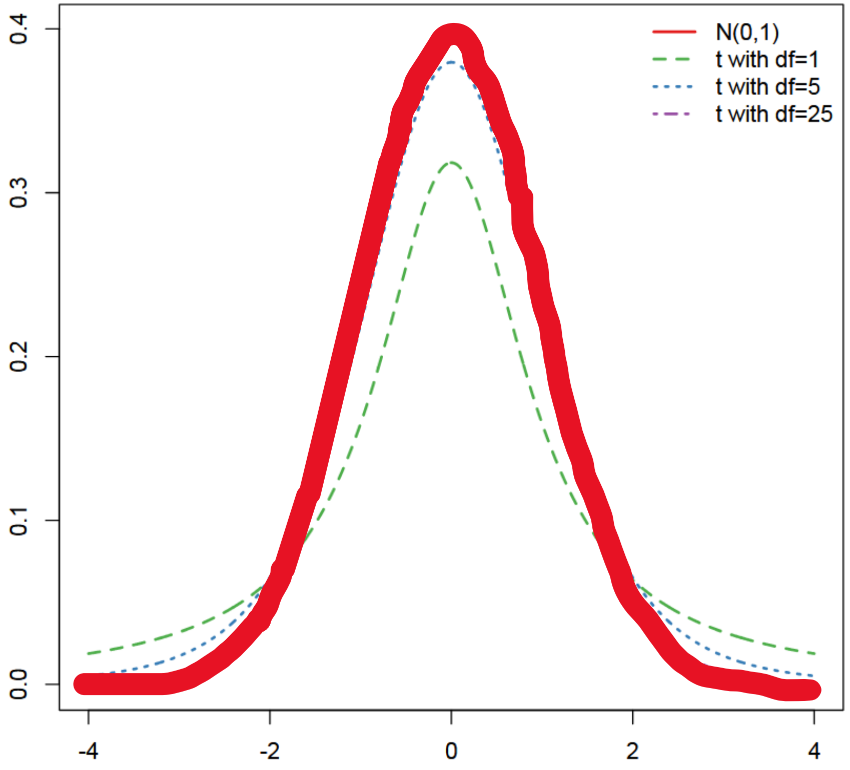


Sampling distribution of $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ is normal with mean 0 and sd 1.

Sampling Distribution

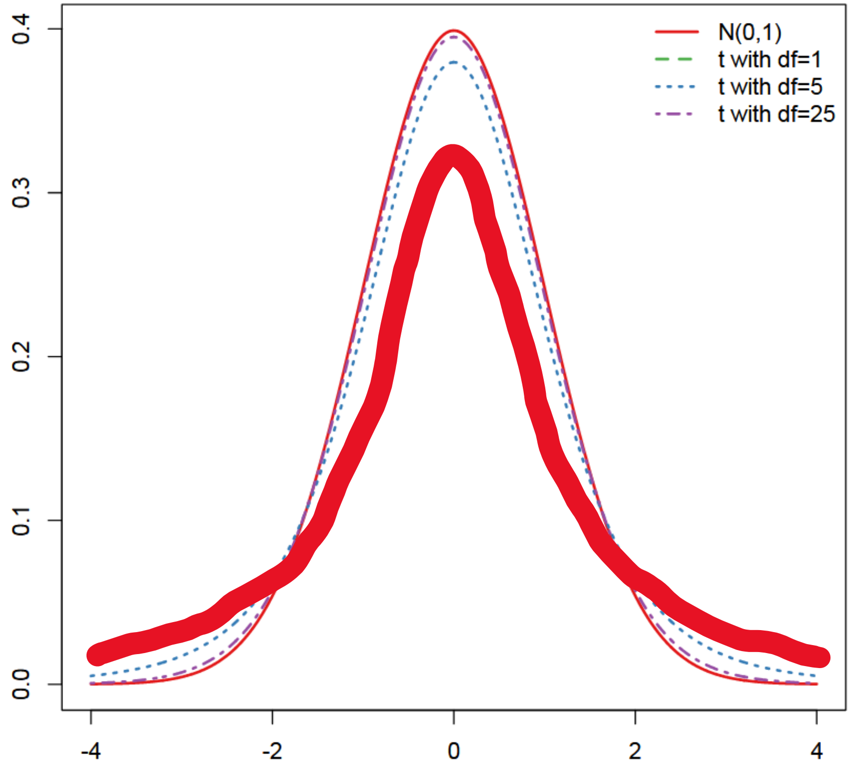
Given pop. distribution with mean μ,
the sampling distribution of size n looks like:

For σ known:



Sampling distribution of $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ is normal with mean 0 and sd 1.

For σ unknown:
Use sample mean s_x to approximate σ.

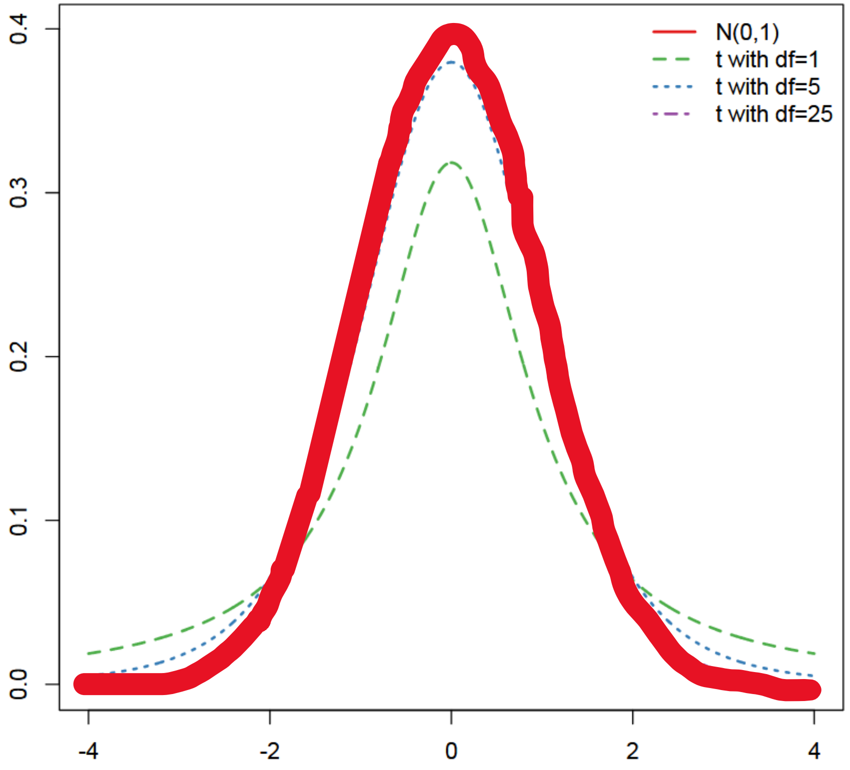


Sampling distribution of $t = \frac{\bar{x}-\mu}{s_x/\sqrt{n}}$ is a t_{n-1}-distribution.

Sampling Distribution

Given pop. distribution with mean μ, the sampling distribution of size n looks like:

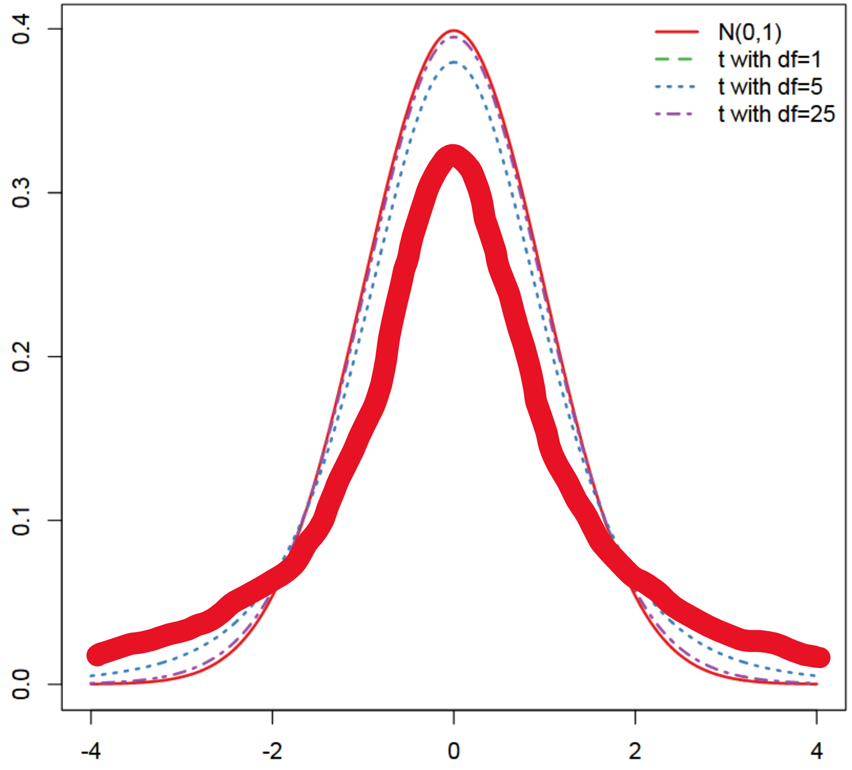
For σ known:



Sampling distribution of $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ is normal with mean 0 and sd 1.

For σ unknown:

Use sample mean s_x to approximate σ.

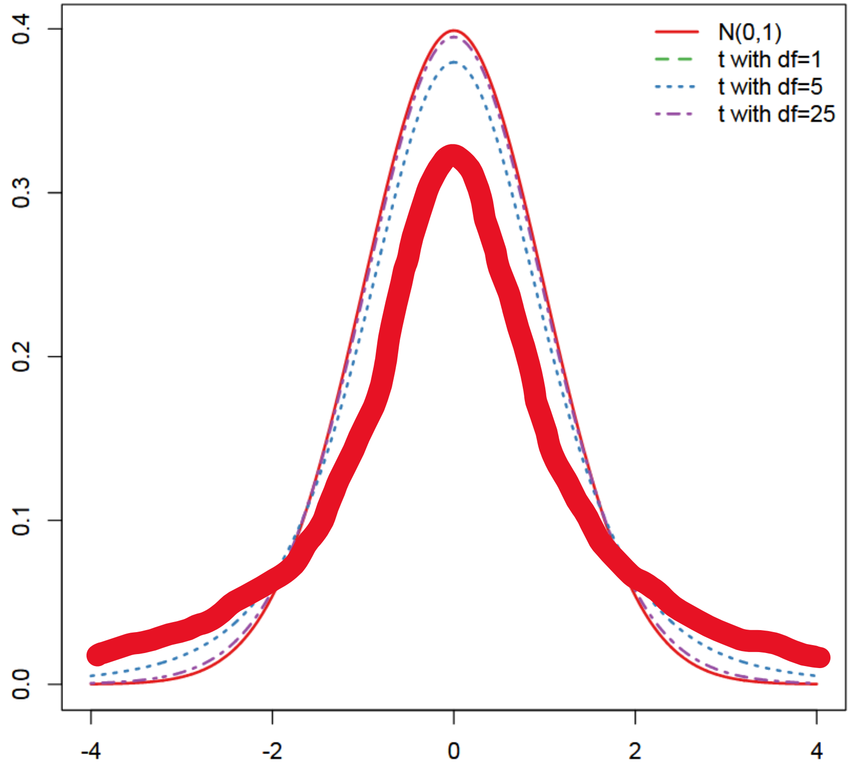


Sampling distribution of $t = \frac{\bar{x}-\mu}{s_x/\sqrt{n}}$ is a t_{n-1}-distribution. The $df = n-1$ is the **degrees of freedom (df)** of the t_{n-1}-distribution.

t_{n-1}-Distribution

Student's t_{n-1}-distribution is a family of symmetric curves fatter than the normal curve.

As sample size n increases, it becomes more like the normal curve.

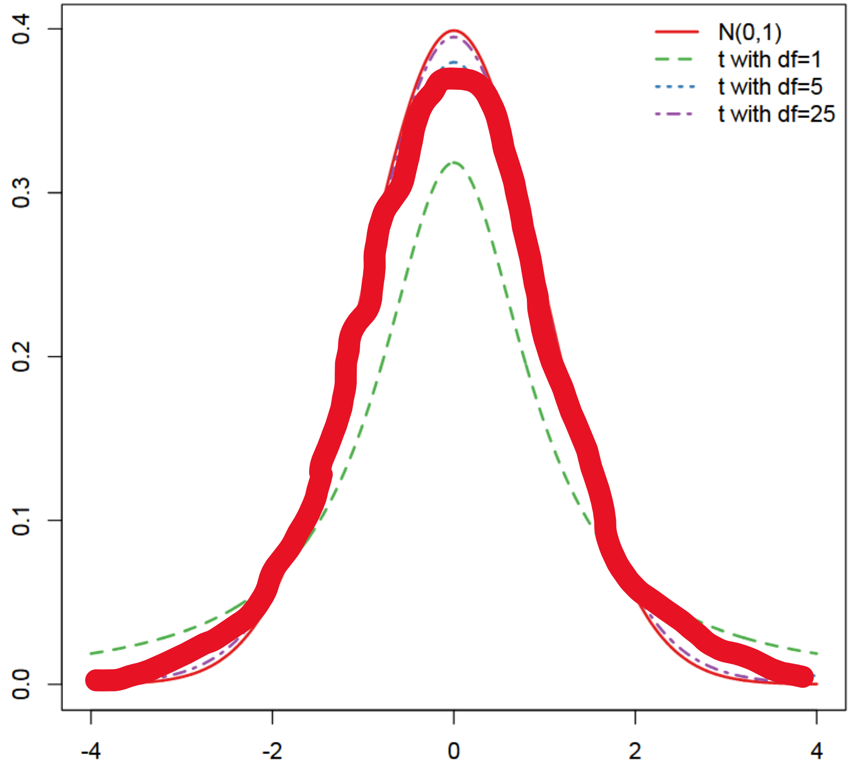


t_1-distribution

t_{n-1}-Distribution

Student's t_{n-1}-distribution is a family of symmetric curves fatter than the normal curve.

As sample size n increases, it becomes more like the normal curve.

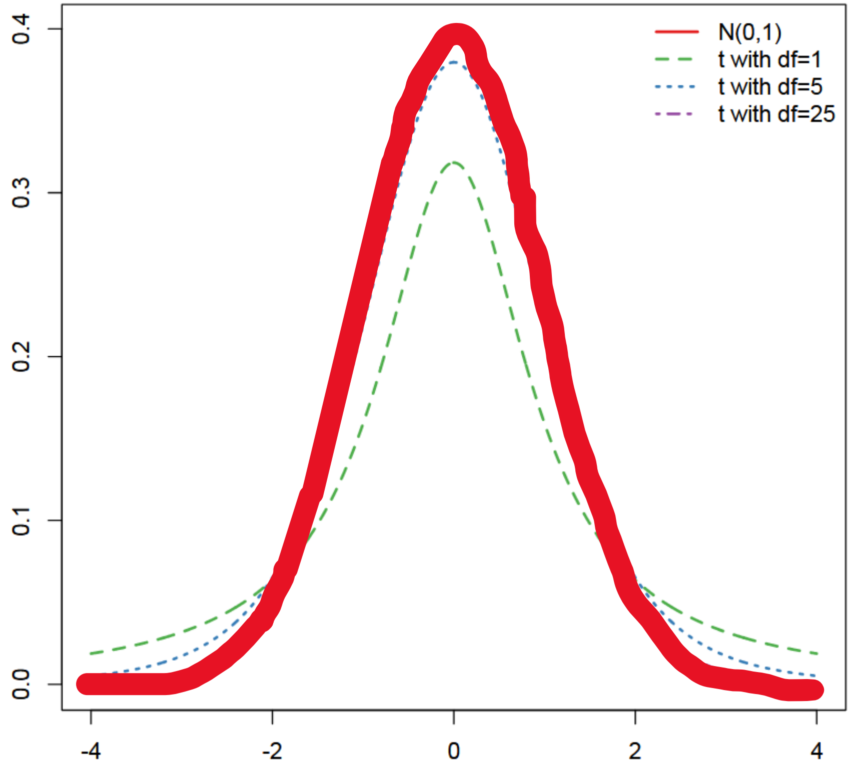


t_5-distribution

t_{n-1}-Distribution

Student's t_{n-1}-distribution is a family of symmetric curves fatter than the normal curve.

As sample size n increases, it becomes more like the normal curve.



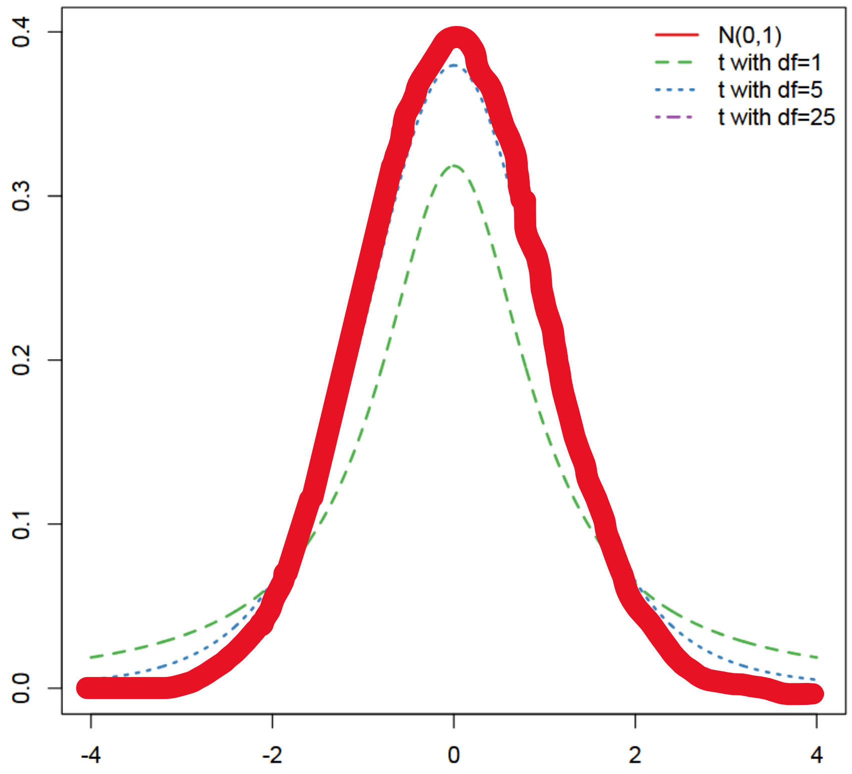
t_{25}-distribution

t_{n-1}-Distribution

Student's t_{n-1}-distribution is a family of symmetric curves fatter than the normal curve.

As sample size n increases, it becomes more like the normal curve.

Gosset wrote internal article "The application of the law of error to work of the brewery" in 1904 for Guinness Brewery, on getting statistical results from small samples.



t_{25}-distribution

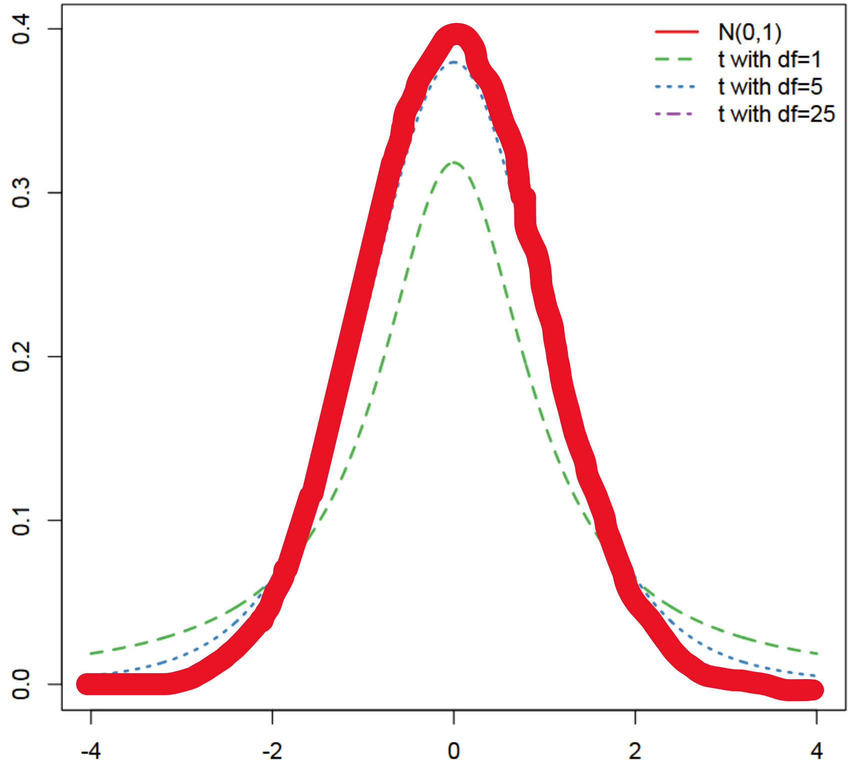
t_{n-1}-Distribution

Student's t_{n-1}-distribution is a family of symmetric curves fatter than the normal curve.

As sample size n increases, it becomes more like the normal curve.

Gosset wrote internal article "The application of the law of error to work of the brewery" in 1904 for Guinness Brewery, on getting statistical results from small samples.

Gosset published externally under the pseudonym **Student**.



t_{25}-distribution

Confidence Interval of Mean with σ unknown (Facts)

The $\boldsymbol{c}$**-confidence interval for population mean** $\boldsymbol{\mu}$ obtained from an SRS with sample mean $\bar{x}$ and sample sd s_x is

$$\bar{x} - E < \mu < \bar{x} + E$$

with **margin of error**

$$E = t_{c,n-1} \frac{s_x}{\sqrt{n}}$$

when the following conditions are met:

- **10% Condition:** $N \geq 10n$.
- **Approx. normal:** Pop. distribution is normal or $n \geq 30$.

Confidence Interval of Mean with σ unknown (Facts)

The $\boldsymbol{c}$**-confidence interval for population mean** $\boldsymbol{\mu}$ obtained from an SRS with sample mean $\bar{x}$ and sample sd s_x is

$$\bar{x} - E < \mu < \bar{x} + E$$

with **margin of error**

$$E = t_{c,n-1} \frac{s_x}{\sqrt{n}}$$

when the following conditions are met:

- **10% Condition:** $N \geq 10n$.
- **Approx. normal:** Pop. distribution is normal or $n \geq 30$.

Compute $t_{c,n-1}$ as **qt((1-c)/2, df = n-1)** in **R.**

Confidence Interval of Mean with σ unknown (Example 1)

The $\boldsymbol{c}$**-confidence interval for population mean** $\boldsymbol{\mu}$ obtained from an SRS with sample mean $\bar{x}$ and sample sd s_x is

$$\bar{x} - E < \mu < \bar{x} + E$$

with **margin of error**

$$E = t_{c,n-1} \frac{s_x}{\sqrt{n}}$$

when the following conditions are met:

- **10% Condition:** $N \geq 10n$.
- **Approx. normal:** Pop. distribution is normal or $n \geq 30$.

Compute $t_{c,n-1}$ as **qt((1-c)/2, df = n-1)** in **R.**

Example 1.
Construct a 95% confidence interval for mean NOX exhaust level in all US cars given a sample of 40 readings with mean 1.268 and sample sd 0.333.

Confidence Interval of Mean with σ unknown (Example 1)

The **c-confidence interval for population mean μ** obtained from an SRS with sample mean $\bar{x}$ and sample sd s_x is

$$\bar{x} - E < \mu < \bar{x} + E$$

with **margin of error**

$$E = t_{c,n-1} \frac{s_x}{\sqrt{n}}$$

when the following conditions are met:

- **10% Condition:** $N \geq 10n$.
- **Approx. normal:** Pop. distribution is normal or $n \geq 30$.

Compute $t_{c,n-1}$ as **qt((1-c)/2, df = n-1)** in **R.**

Example 1.
Construct a 95% confidence interval for mean NOX exhaust level in all US cars given a sample of 40 readings with mean 1.268 and sample sd 0.333.

Answer.
$N \geq 10 \cdot 40$ and $n \geq 30$ so we can proceed.

Confidence Interval of Mean with σ unknown (Example 1)

The $\boldsymbol{c}$**-confidence interval for population mean** $\boldsymbol{\mu}$ obtained from an SRS with sample mean $\bar{x}$ and sample sd s_x is

$$\bar{x} - E < \mu < \bar{x} + E$$

with **margin of error**

$$E = t_{c,n-1} \frac{s_x}{\sqrt{n}}$$

when the following conditions are met:

- **10% Condition:** $N \geq 10n$.
- **Approx. normal:** Pop. distribution is normal or $n \geq 30$.

Compute $t_{c,n-1}$ as **qt((1-c)/2, df = n-1)** in **R.**

Example 1.
Construct a 95% confidence interval for mean NOX exhaust level in all US cars given a sample of 40 readings with mean 1.268 and sample sd 0.333.

Answer.
$N \geq 10 \cdot 40$ and $n \geq 30$ so we can proceed.

$c = 95\% \Rightarrow t_{c,n-1} =$ **qt(0.025, df = 39) = -2.023**

Confidence Interval of Mean with σ unknown (Example 1)

The $\boldsymbol{c}$**-confidence interval for population mean** $\boldsymbol{\mu}$ obtained from an SRS with sample mean $\bar{x}$ and sample sd s_x is

$$\bar{x} - E < \mu < \bar{x} + E$$

with **margin of error**

$$E = t_{c,n-1} \frac{s_x}{\sqrt{n}}$$

when the following conditions are met:

- **10% Condition:** $N \geq 10n$.
- **Approx. normal:** Pop. distribution is normal or $n \geq 30$.

Compute $t_{c,n-1}$ as **qt((1-c)/2, df = n-1)** in **R.**

Example 1.
Construct a 95% confidence interval for mean NOX exhaust level in all US cars given a sample of 40 readings with mean 1.268 and sample sd 0.333.

Answer.
$N \geq 10 \cdot 40$ and $n \geq 30$ so we can proceed.

$c = 95\% \Rightarrow t_{c,n-1} =$ **qt(0.025, df = 39) = -2.023**

$$E = 2.023 \frac{0.333}{\sqrt{40}} = 0.107$$

Confidence Interval of Mean with σ unknown (Example 1)

The $\boldsymbol{c}$**-confidence interval for population mean** $\boldsymbol{\mu}$ obtained from an SRS with sample mean $\bar{x}$ and sample sd s_x is

$$\bar{x} - E < \mu < \bar{x} + E$$

with **margin of error**

$$E = t_{c,n-1} \frac{s_x}{\sqrt{n}}$$

when the following conditions are met:

- **10% Condition:** $N \geq 10n$.
- **Approx. normal:** Pop. distribution is normal or $n \geq 30$.

Compute $t_{c,n-1}$ as **qt((1-c)/2, df = n-1)** in **R.**

Example 1.
Construct a 95% confidence interval for mean NOX exhaust level in all US cars given a sample of 40 readings with mean 1.268 and sample sd 0.333.

Answer.
$N \geq 10 \cdot 40$ and $n \geq 30$ so we can proceed.

$c = 95\% \Rightarrow t_{c,n-1} =$ **qt(0.025, df = 39) = -2.023**

$$E = 2.023 \frac{0.333}{\sqrt{40}} = 0.107$$

$$1.268 - 0.107 < \mu < 1.268 + 0.107$$

Confidence Interval of Mean with σ unknown (Example 1)

The $\boldsymbol{c}$**-confidence interval for population mean** $\boldsymbol{\mu}$ obtained from an SRS with sample mean $\bar{x}$ and sample sd s_x is

$$\bar{x} - E < \mu < \bar{x} + E$$

with **margin of error**

$$E = t_{c,n-1} \frac{s_x}{\sqrt{n}}$$

when the following conditions are met:

- **10% Condition:** $N \geq 10n$.
- **Approx. normal:** Pop. distribution is normal or $n \geq 30$.

Compute $t_{c,n-1}$ as **qt((1-c)/2, df = n-1)** in **R.**

Example 1.
Construct a 95% confidence interval for mean NOX exhaust level in all US cars given a sample of 40 readings with mean 1.268 and sample sd 0.333.

Answer.
$N \geq 10 \cdot 40$ and $n \geq 30$ so we can proceed.

$c = 95\% \Rightarrow t_{c,n-1} =$ **qt(0.025, df = 39) = -2.023**

$$E = 2.023 \frac{0.333}{\sqrt{40}} = 0.107$$

$$1.268 - 0.107 < \mu < 1.268 + 0.107$$

Conclusion: We are 95% confident that the true mean NOX exhaust level is between 1.161 and 1.375.