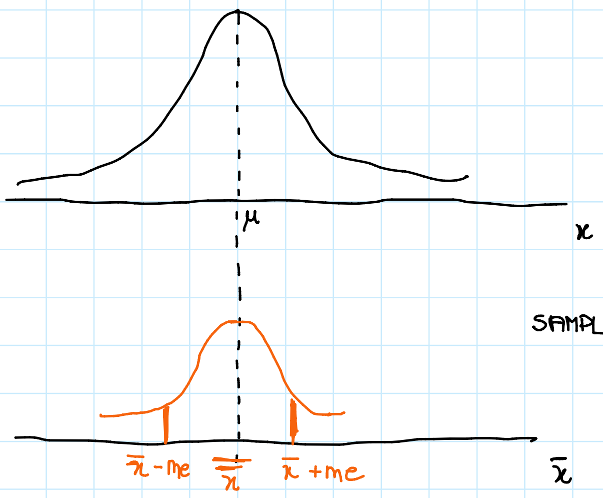
Confidence intervals

martes, 9 de agosto de 2022 12:50 p. m.

PARENT DISTRIBUTION

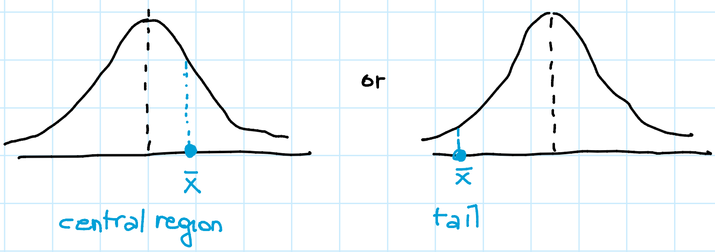


SAMPLE DISTRIBUTION

$\bar{x} \to [\bar{x}-me, \bar{x}+me]$

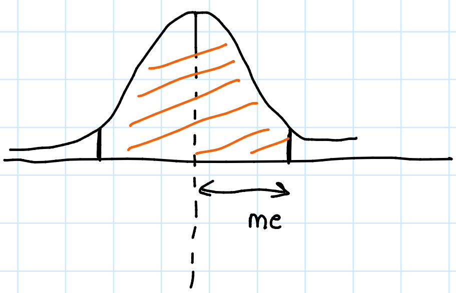
me = margin of error

Choosing a random sample and measuring $\bar{x}$, $\bar{x}$ can fall at any point of the sample distribution:



extreme values are less frequent, but possible.

Sampling distribution of $\bar{x}$ is normal with standard deviation $\frac{\sigma}{\sqrt{N}}$ (standard error)



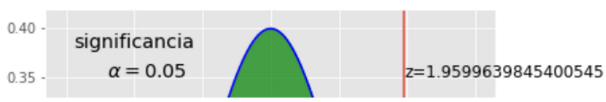
One can choose the margin of error such that the area under the curve centered around the mean corresponds to some "confidence level" (typically 95%, 90%, 99%)

The margin of error is expressed as a multiple of the standard error: $z^* \cdot \frac{\sigma}{\sqrt{N}}$ the value of the constant z^* depends on the confidence level and corresponds to the quantiles of the standard normal distribution corresponding to the complement of the confidence level (significance level)

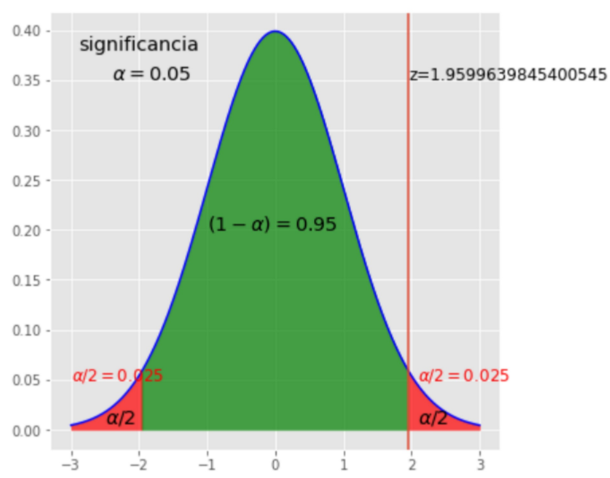
Example: 95%

significance level = α = 1 − conf level

[ ]



$\alpha = 1-0.95 = 0.05$



$\alpha = 1 - 0.95 = 0.05$

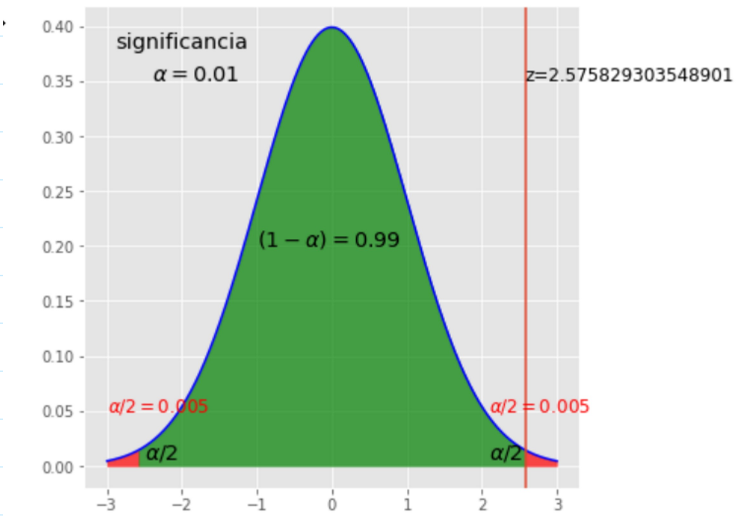
The significance is equally distributed on the tails

$\frac{\alpha}{2} = 0.025$

z^* corresponds to the quantile $P(z \le z^*) = 0.975$

or $P(z \le z^*) = 0.025$

Example: 99%



$\alpha = 1 - 0.99 = 0.01$

$\alpha/2 = 0.005$

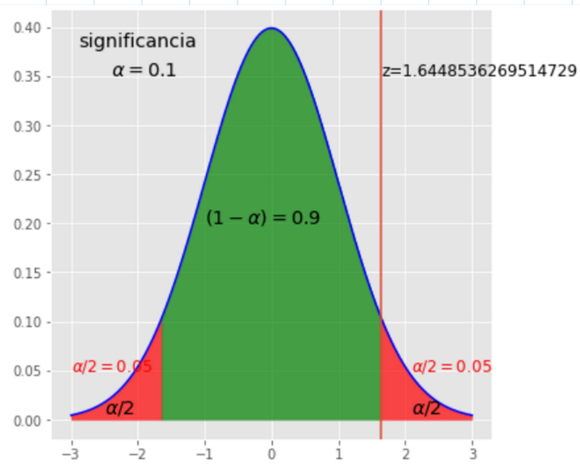
z^* is such that

$P(z \le z^*) = 0.995$ (right tail)

or

$P(z \le z^*) = 0.005$ (left tail)

Example 90%



$\alpha = 1 - 0.9 = 0.1$

$\alpha/2 = 0.05$

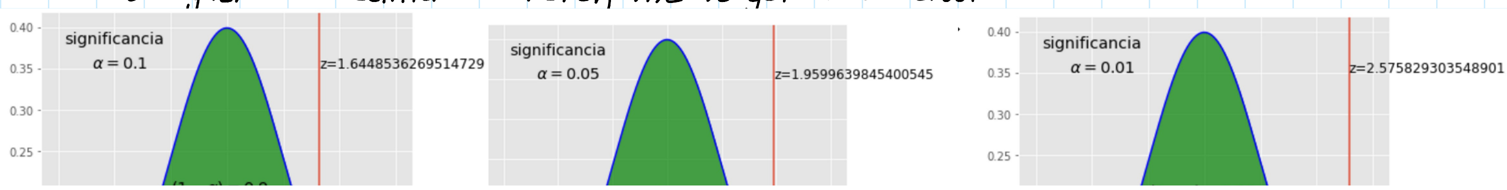
z^* is such that:

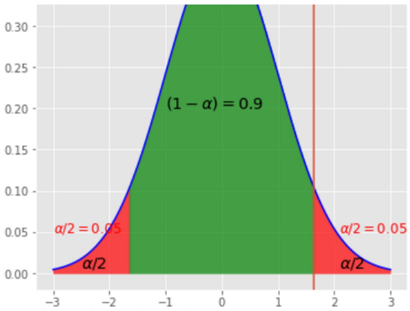
$P(z \le z^*) = 0.95$ (right tail)

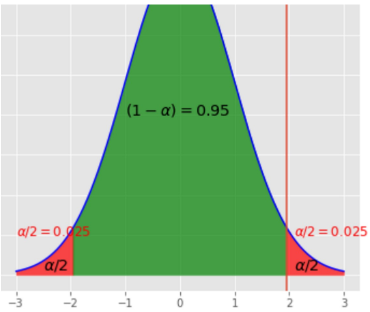
or

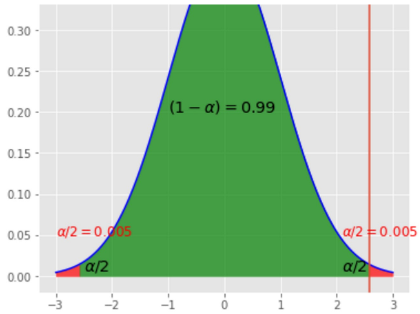
$P(z \le z^*) = 0.05$ (left tail)

The higher the confidence level, the larger the interval:









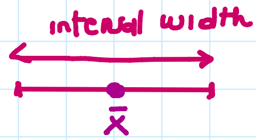
estimate more precise 90%

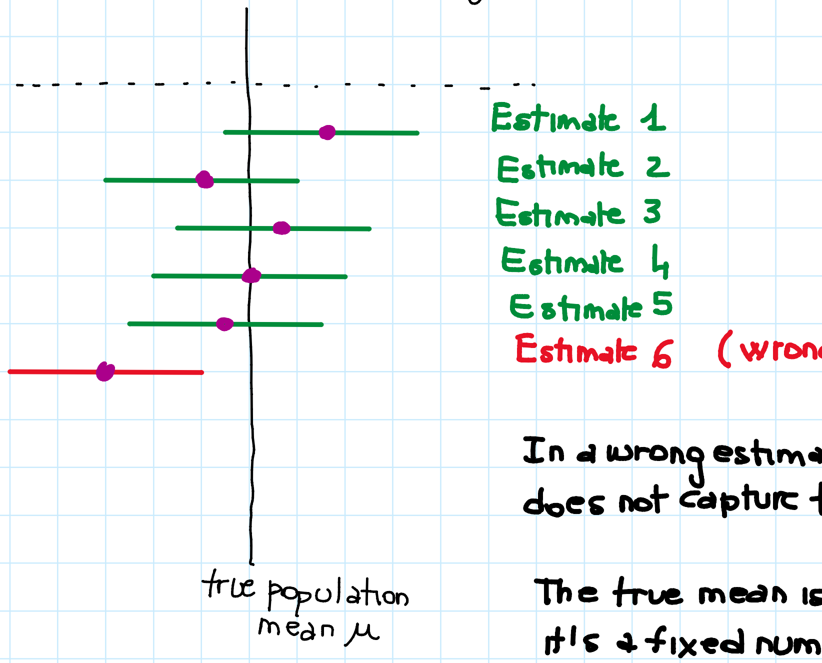
95%

estimate less precise 99%

Interpretation of confidence intervals :

Let's suppose that I take 100 different samples with the same size N and make the estimate of the interval 100 times, choosing a confidence level of 95% :





In a wrong estimate, the interval does not capture the true mean.

The true mean is not a random variable, it's a fixed number. The randomness is in the sample choice.