Probabilidade e Estatística

Teorema Central do Limite e Intervalo de Confiança

Teorema Central do Limite

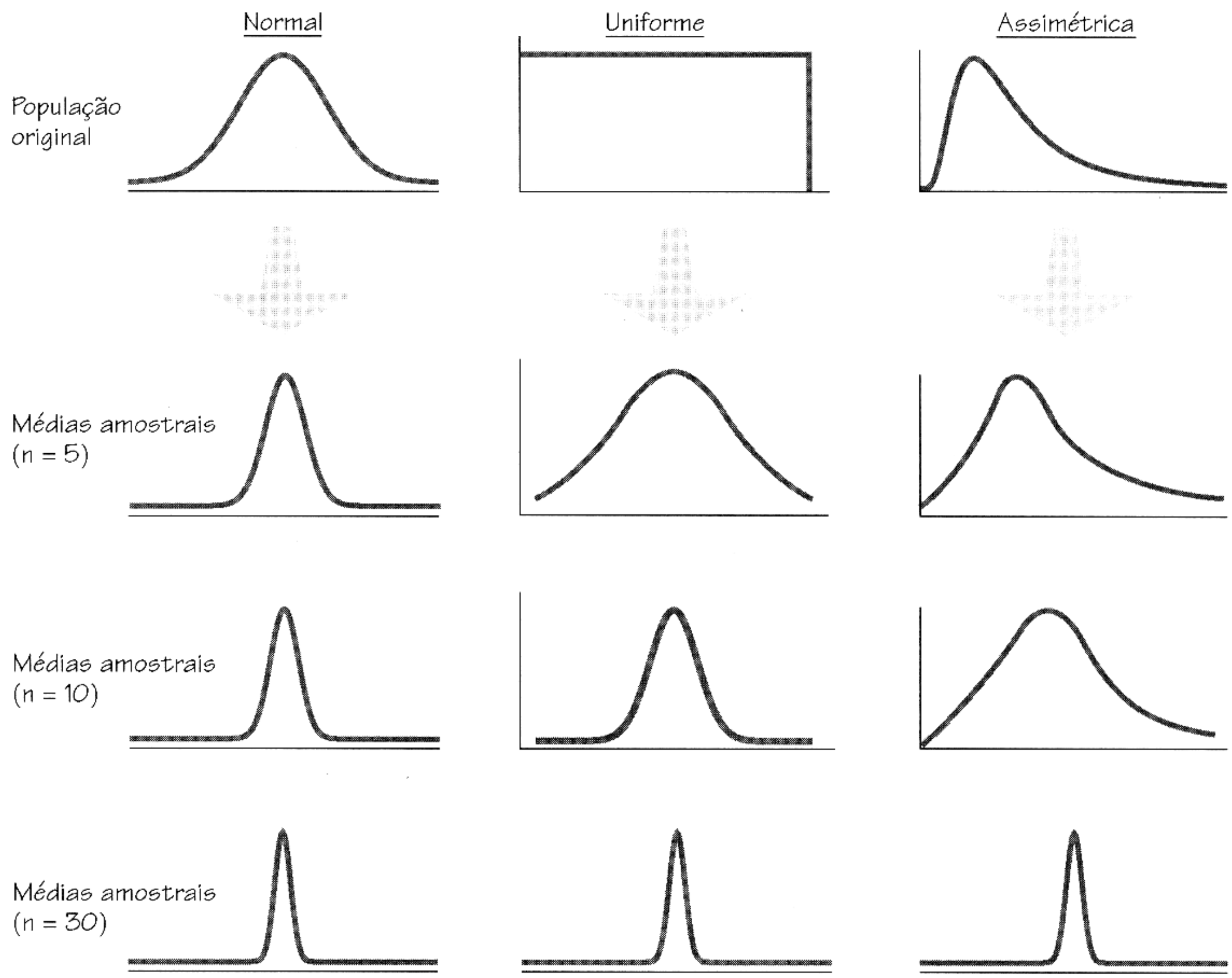


Fig. 5-21 Distribuição normal, uniforme e assimétrica.

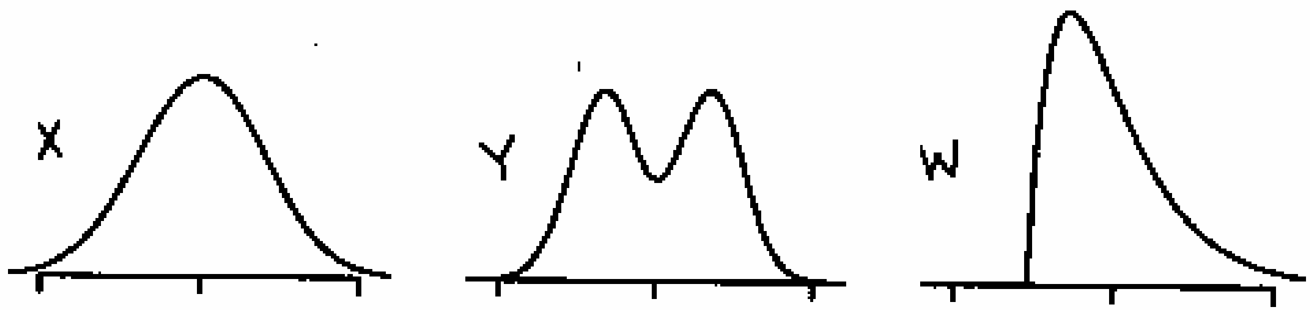
Teorema Central do Limite

- Um variável aleatória pode ter uma distribuição qualquer (normal, uniforme,...), possuindo uma média μ e um desvio-padrão σ.
- Se, ao invés de tirarmos uma única amostra (digamos, 100 coletas), tirarmos várias amostras de tamanho ´n´ (digamos, 20 amostras compostas por cinco coletas: 20x5=100 coletas) e analisarmos a distribuição das médias de cada amostra de tamanho ´n´, observaremos que:
  - À medida que o tamanho ´n´ da amostra aumenta, a distribuição das médias amostrais tende a uma distribuição normal
  - A média das médias amostrais será a média populacional $\mu = \bar{\bar{x}} = \dfrac{\sum \bar{x}}{m}$
  - O desvio padrão das médias amostrais será $\sigma / \sqrt{n}$

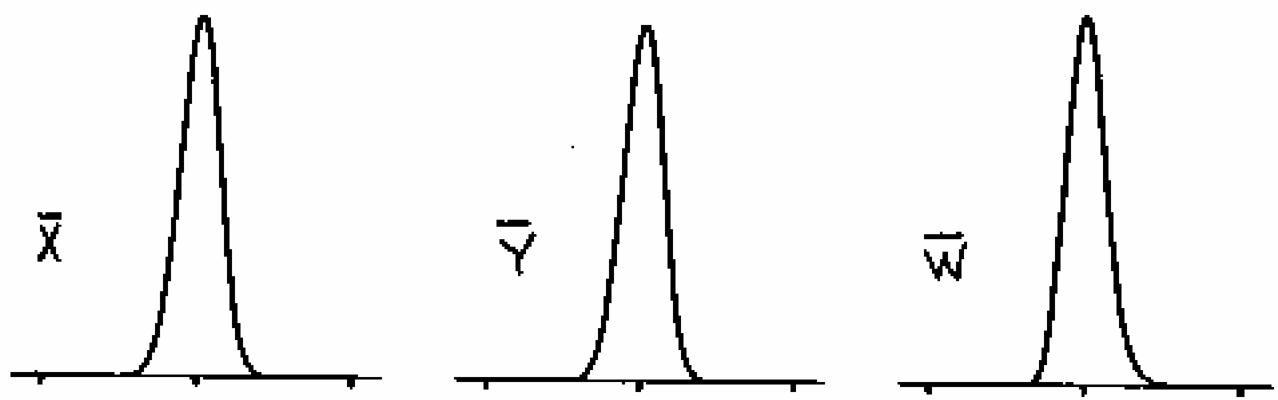
Teorema Central do Limite

- Observações importantes:
  - Quando maior o tamanho das amostras, a distribuição das médias será mais próxima de uma distribuição normal.
  - Para n>30, a distribuição das médias amostrais pode ser aproximada satisfatoriamente por uma distribuição normal.
  - Se a distribuição da variável ‘x’ for originalmente uma distribuição normal, então a distribuição das médias amostrais terá distribuição normal para qualquer tamanho amostral ‘n’.

O QUE HÁ DE EXTRAORDINÁRIO NO **TEOREMA DO LIMITE CENTRAL**? ELE NOS DIZ QUE QUALQUER QUE SEJA A FORMA DA DISTRIBUIÇÃO ORIGINAL, SUAS **MÉDIAS** RESULTAM NUMA DISTRIBUIÇÃO **NORMAL.** PARA ENCONTRARMOS A DISTRIBUIÇÃO DA MÉDIA, BASTA CONHECERMOS A MÉDIA DA POPULAÇÃO E O DESVIO PADRÃO.



TODAS AS TRÊS DENSIDADES ACIMA TÊM A MESMA MÉDIA E DESVIO PADRÃO. APESAR DE SUAS FORMAS DIFERENTES, QUANDO n=10, AS DISTRIBUIÇÕES DAS MÉDIAS DAS AMOSTRAS SÃO PRATICAMENTE IDÊNTICAS.



Estimativa de Média Populacional

- Supondo que coletemos 20 amostras de alturas de alunos e considerando que esta representa efetivamente a população de alunos da universidade.
- Como estimativa da média da população (μ) de alunos, poderíamos utilizar:
  - A média
  - A moda
  - A mediana
  - Ponto médio

Estimativa de Média Populacional

- Em geral, entretanto, a média amostral $\bar{x}$ do conjunto de dados é a melhor estimativa de uma média populacional.
- Uma **estimativa** é um valor específico, ou um intervalo de valores usados para aproximar um parâmetro populacional.
- Um **estimador** é uma característica da amostra (Ex: $\bar{x}$), utilizado para obtermos uma aproximação do parâmetro populacional.

Estimativa de Média Populacional

- Razões para utilizarmos a média amostral como um estimador de uma média populacional μ.
  - A distribuição das médias amostrais $\overline{X}$ tende a apresentar menor variação do que distribuições de outras características amostrais (mediana ou moda)
  - É um estimador não tendencioso da média populacional μ: tende a centrar-se em torno de μ; tende a um valor central que é o próprio valor de μ

Estimativa de Média Populacional

- Como a média amostral é um valor pontual, chamamos a este de estimador pontual.
- Portanto, a média amostral $\bar{x}$ é a melhor estimativa pontual da média populacional μ.
- No nosso exemplo, a suposição da média amostral $\bar{x}$ das 20 amostras é a melhor estimativa pontual da população de alunos da universidade.
- Entretanto,....

Estimativa de Média Populacional

- O que nos garante que as 20 amostras compõem uma boa estimativa da população?
- Por isso associamos uma estimativa pontual a uma outra estimativa:

INTERVALO DE CONFIANÇA

ou

ESTIMATIVA INTERVALAR

Intervalo de Confiança

- É uma amplitude (ou um intervalo) de valores que tem a probabilidade de conter o valor verdadeiro da população
- Observa-se que, na definição de intervalo de confiança, está associado uma probabilidade.
- A esta probabilidade chamamos de:

Nível de Confiança

Grau de Confiança, ou

Coeficiente de Confiança

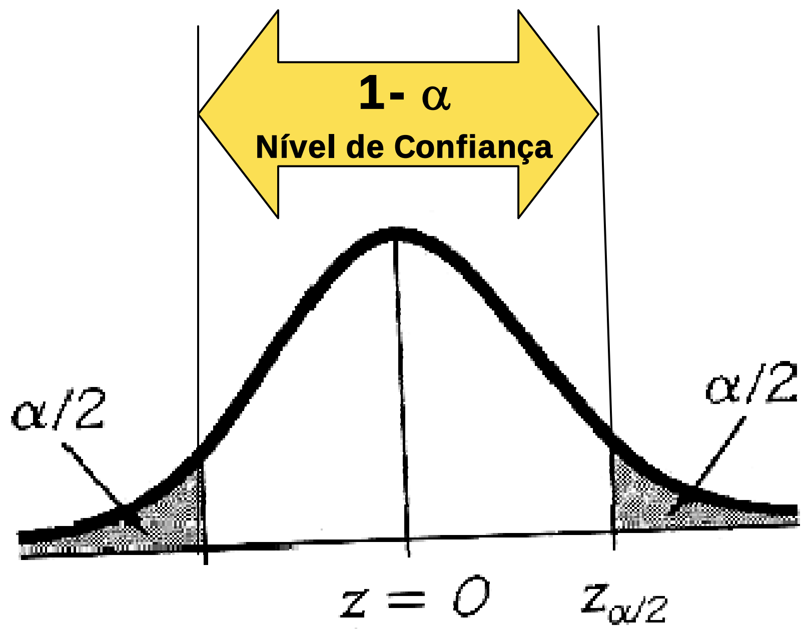
Intervalo de Confiança

$$\text{Probabilidade}\{c_1 \leq \mu \leq c_2\} = 1 - \alpha$$

- **O intervalo (c1,c2) é chamado de <u>intervalo de confiança</u> da média da população.**
- **α é o <u>nível de significância</u>.**
- **100(1- α) é o <u>nível de confiança</u>.**
- **1- α é o <u>coeficiente de confiança</u>.**

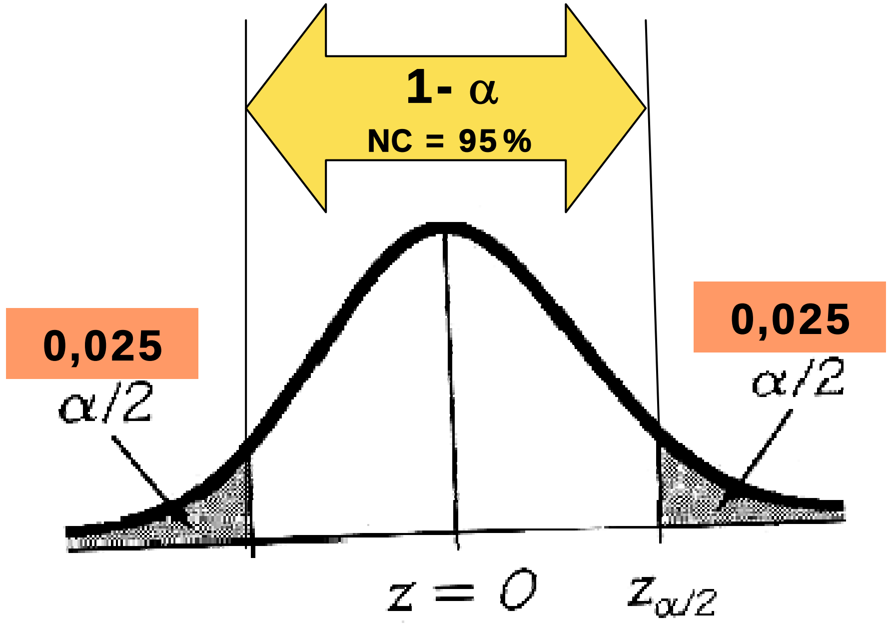
Nível de Confiança (NC)

- É a probabilidade **1-α** (comumente expressa percentualmente) do intervalo de confiança conter o valor verdadeiro, o parâmetro populacional



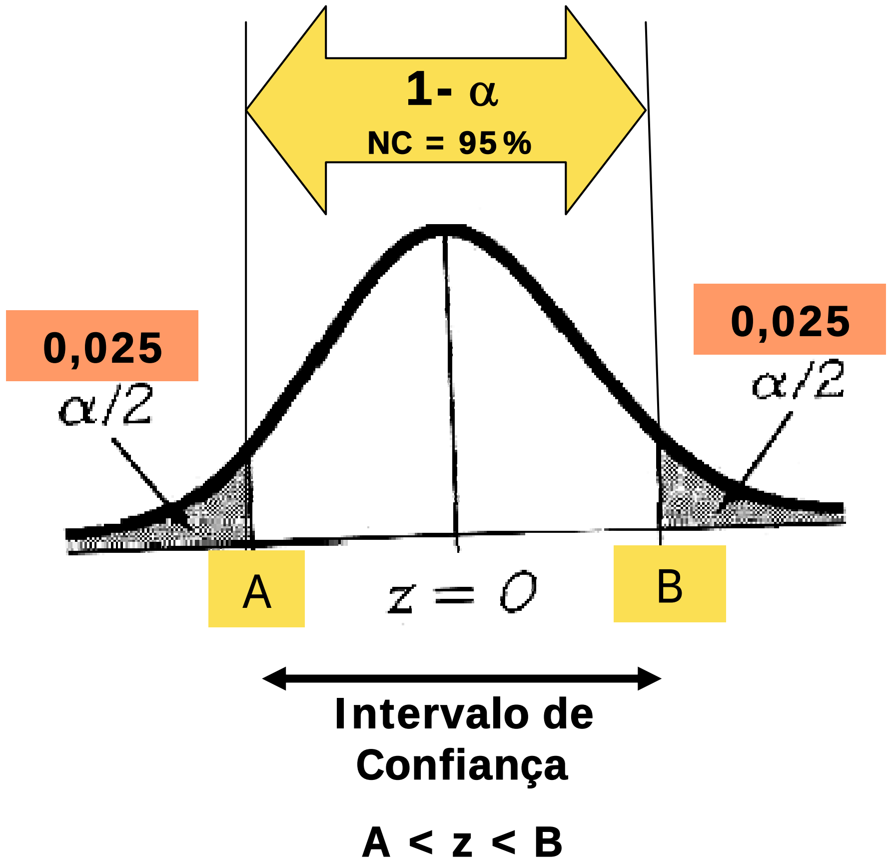
Nível de Confiança (NC)

- Comumente utiliza-se NC de:
- 90% → $\alpha = 0,1$
- 95% → $\alpha = 0,05$
- 99% → $\alpha = 0,01$



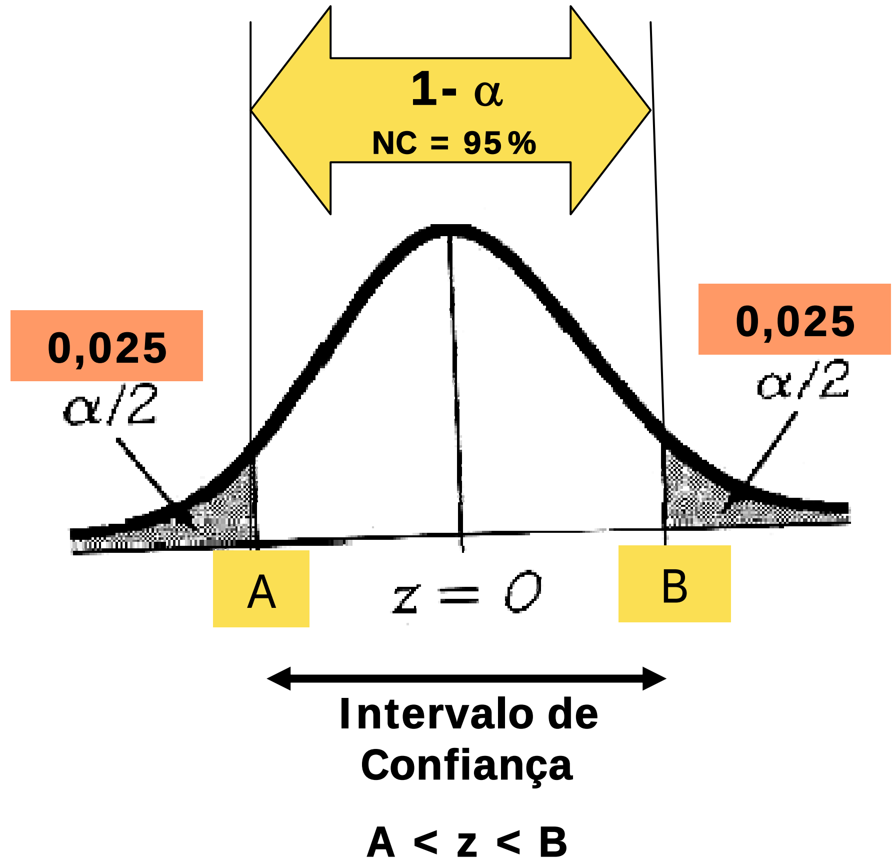
Intervalo de Confiança

- Observações:
  - O Intervalo de Confiança consiste em um intervalo na escala z e está associado a um NC.

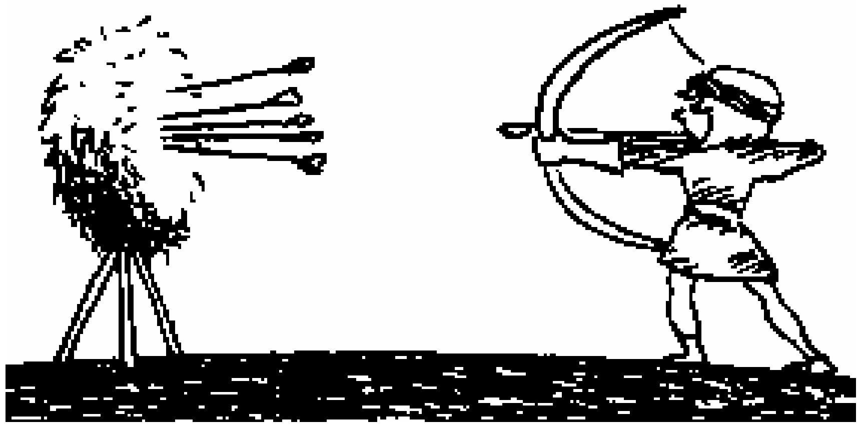


Intervalo de Confiança

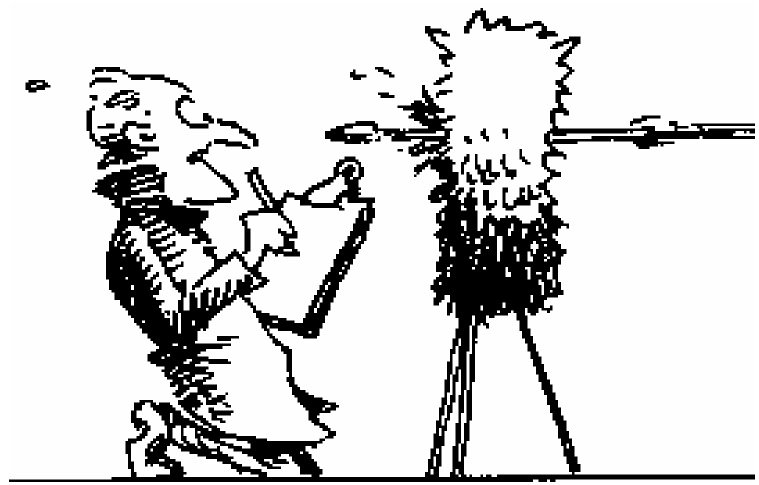
- Conclusão:
  - Se coletarmos várias amostras de 20 alunos e construirmos um intervalo de 95% de confiança para cada uma, a longo prazo, 95% destes intervalos conteriam efetivamente a média da população μ



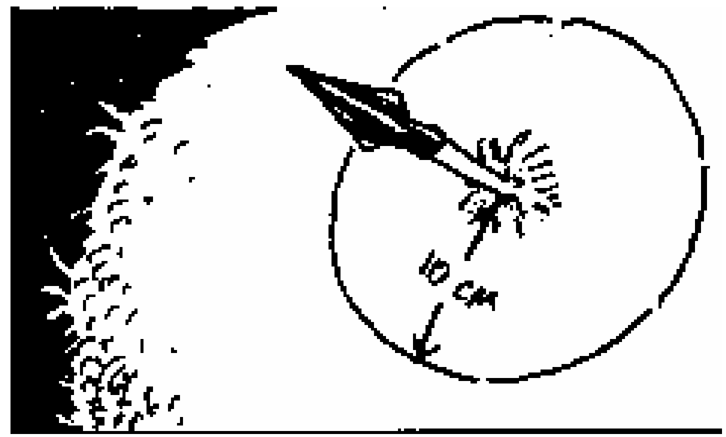
CONSIDERE UMA ARQUEIRA ATIRANDO EM UM ALVO. SUPONHA QUE ELA ACERTA NO CENTRO COM RAIO DE 10 CM 95% DAS VEZES. OU SEJA, ERRA APENAS UMA VEZ A CADA 20 TENTATIVAS.



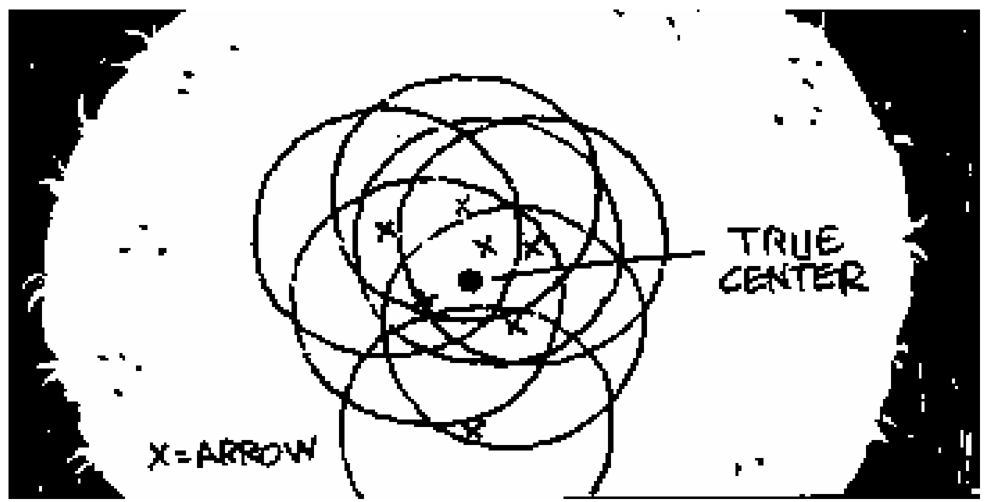
SENTADO ATRÁS DO ALVO ENCONTRA-SE UM BRAVO DETETIVE, QUE NÃO VÊ ONDE ESTÁ O CENTRO. A ARQUEIRA ATIRA A PRIMEIRA FLECHA..



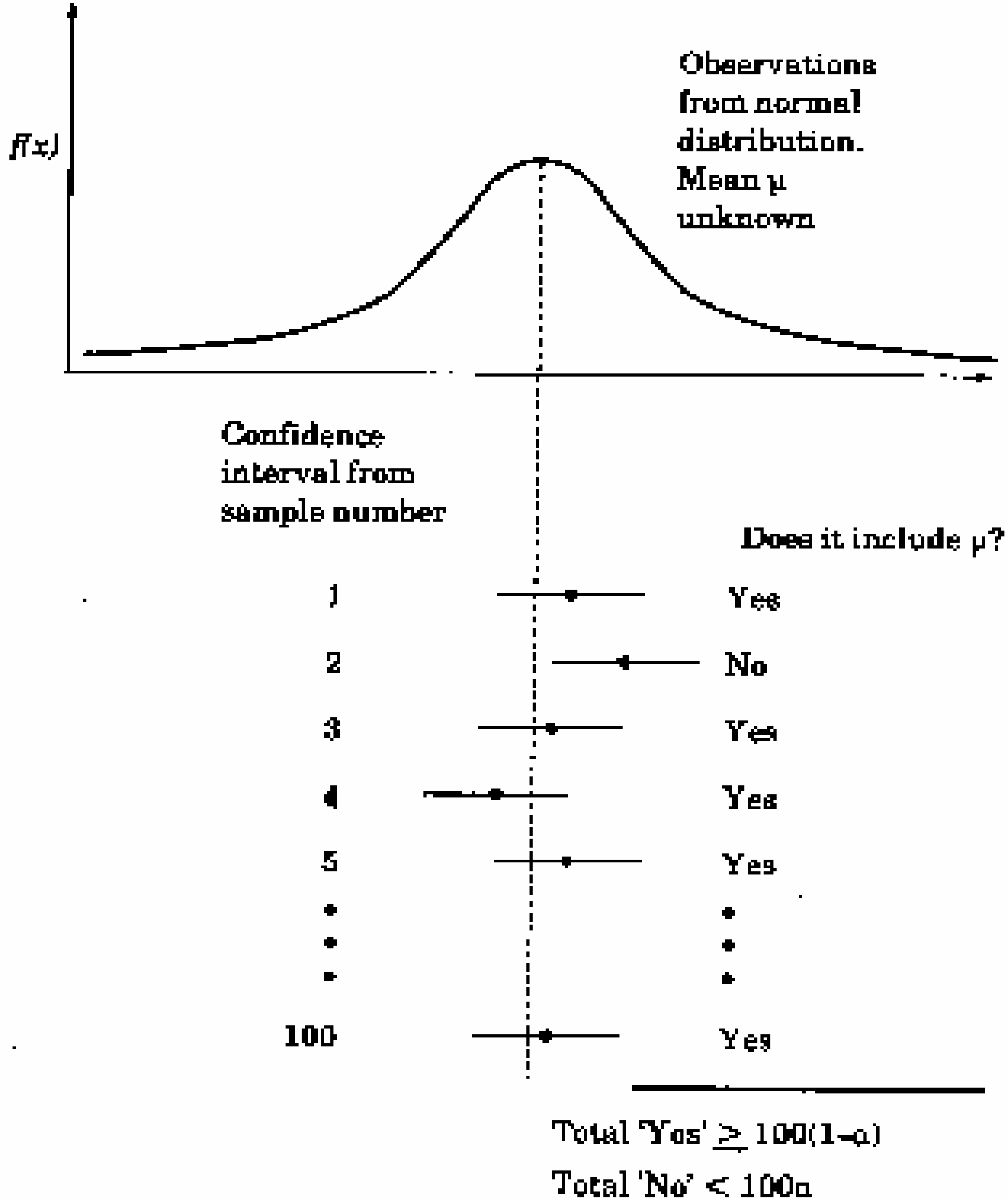
CONHECENDO O NÍVEL DA HABILIDADE DA ARQUEIRA, O DETETIVE DESENHA UM CÍRCULO COM 10 CM DE RAIO AO REDOR DA FLECHA. ELE TEM **95% DE CONFIANÇA** DE QUE O SEU CÍRCULO INCLUI O CENTRO DO ALVO!



ELE RACIOCINOU QUE SE DESENHASSE CÍRCULOS COM 10 CM DE RAIO AO REDOR DE **MUITAS FLECHAS**, OS SEUS CÍRCULOS INCLUIRIAM O CENTRO DO ALVO EM 95% DOS CASOS..



Significado do I.C.



Como melhorar a confiança?

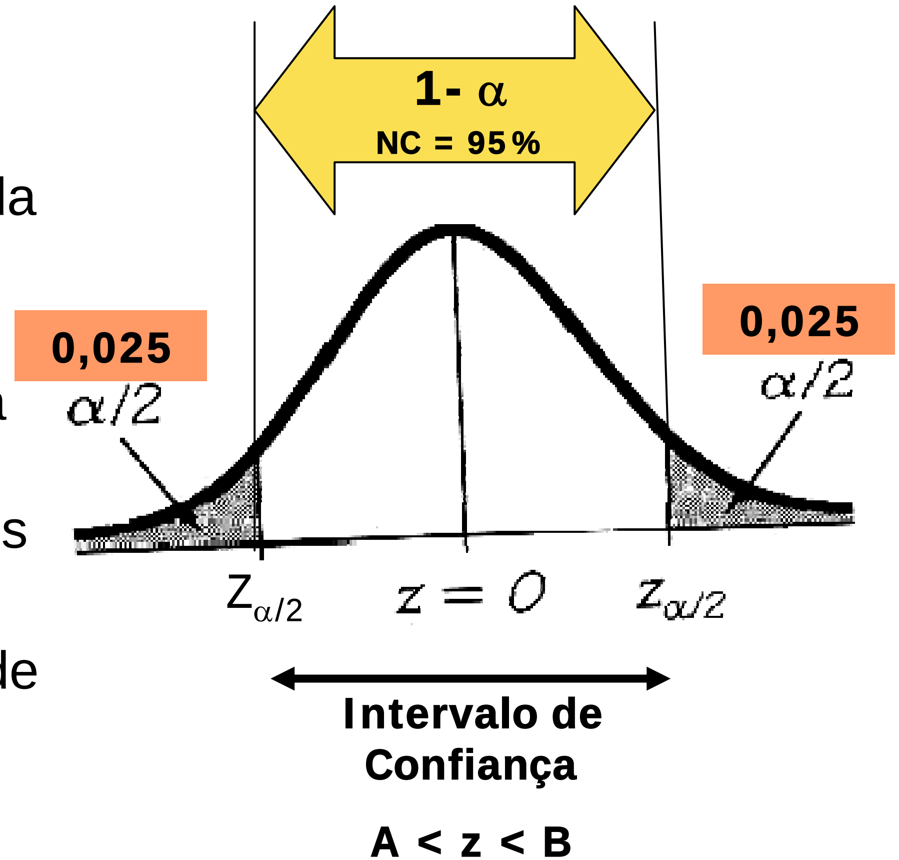
AUMENTANDO O
TAMANHO DO CÍRCULO

OU, MELHORANDO
A MIRA DA ARQUEIRA!

O PRIMEIRO MÉTODO É EQUIVALENTE A **ALARGAR O INTERVALO DE CONFIANÇA**. QUANTO MAIOR FOR A MARGEM DE ERRO, MAIS CERTO VOCÊ ESTÁ DE QUE O VALOR DESEJADO ENCONTRA-SE NO INTERVALO:

Intervalo de Confiança

- Valor Crítico: $Z_{\alpha/2}$
  - Corresponde ao valor de fronteira da área de $\alpha/2$ na cauda direita da distribuição normal padronizada.
  - É o número na fronteira que separa os valores estatísticos amostrais prováveis de ocorrerem, dos valores que tem pouca chance de ocorrer.
  - É um escore *z* com a propriedade de separar uma área de $\alpha/2$ na cauda direita da distribuição normal padronizada



Observação Importante

- Pelo Teorema Central do Limite sabemos que as médias amostrais $\overline{x}$ tendem a distribuir-se por uma normal. Assim, a área sombreada apresenta chance relativamente pequena de conter uma média amostral.
- Denotando de $\alpha/2$ a área sombreada de cada extremo, há uma probabilidade de α da média amostral estar em um dos extremos. Pela regra do complemento há uma probabilidade de $1 - \alpha$ da média amostral estar na região não sombreada.

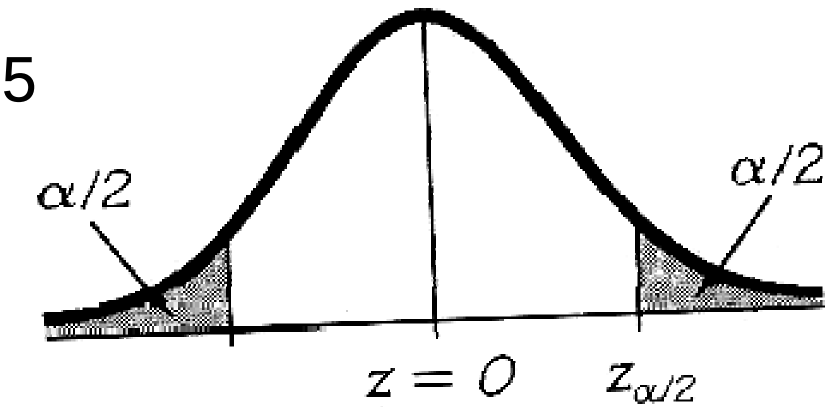
Exercício

- Calcule o valor crítico $Z_{\alpha/2}$ que corresponde ao NC de 90%.

NC = 0,90 => α=0,10 => $\alpha/2$=0,05

Na tabela de Distribuição Normal

- $\alpha/2 = 0,05$
- Área entre Z=0 e Z=$\alpha/2$ é 0,45
- $Z_{\alpha/2} = 1,645$



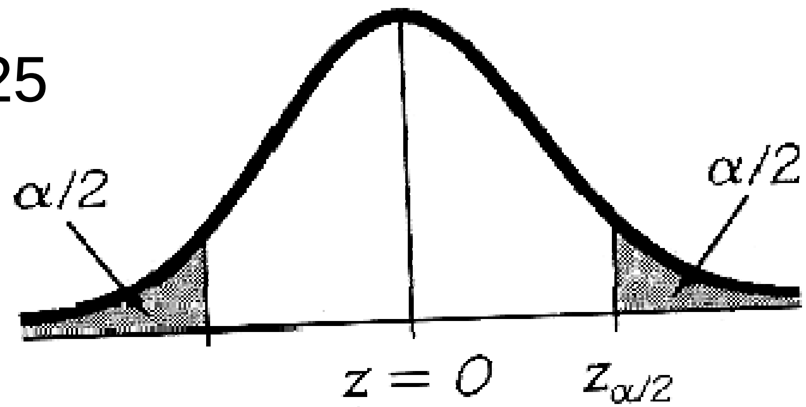
Exercício

- Calcule o valor crítico $Z_{\alpha/2}$ que corresponde ao NC de 95%.

NC = 0,95 => α=0,05 => $\alpha/2$=0,025

Na tabela de Distribuição Normal

- $\alpha/2 = 0,025$
- Área entre Z=0 e Z=$\alpha/2$ é 0,475
- $Z_{\alpha/2} = 1,96$



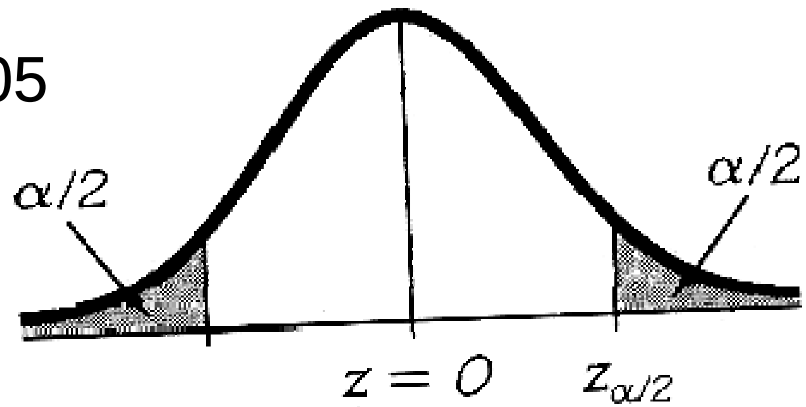
Exercício

- Calcule o valor crítico $Z_{\alpha/2}$ que corresponde ao NC de 99%.

NC = 0,99 => α=0,01 => $\alpha/2$=0,005

Na tabela de Distribuição Normal

- $\alpha/2$ = 0,005
- Área entre Z=0 e Z=$\alpha/2$ é 0,495
- $Z_{\alpha/2}$ = 2,575



Exercícios

- Uma fábrica de parafusos tem por especificação fabricá-los com diâmetro médio de 25mm. Para controle de seu processo, coleta-se ao longo de um dia de trabalho 35 amostras de tamanho 30 (n=30). A média das médias das amostras e o desvio padrão de um determinado dia acusaram:

$$\bar{\bar{x}} = 25mm \qquad \bar{s} = 1{,}5mm$$

- Calcule o diâmetro inferior e o superior representativo dos valores críticos $Z_{\alpha/2}$ desta distribuição para NC = 99%.

Resposta

$$z_{\alpha/2} = \frac{x - \overline{\overline{x}}}{\overline{s}}$$

$$\Rightarrow \quad 2{,}575 = \frac{x - 25}{1{,}5} \Rightarrow 3{,}8625 = x - 25 \quad \Rightarrow \quad x_s = 28{,}86mm$$

$$\Rightarrow \quad -2{,}575 = \frac{x - 25}{1{,}5} \Rightarrow -3{,}8625 = x - 25 \quad \Rightarrow \quad x_i = 21{,}14mm$$

- O que isto significa?
  - Existe 99% de probabilidade do intervalo de 21,14 e 28,86mm conter a média populacional de diâmetro de parafuso, ou
  - A fábrica possui 99% de chance de produzir lotes de peças com médias entre 21,14 e 28,86mm.