Probabilidade e Estatística

Teorema do Limite Central e Intervalo de Confiança

Teorema do Limite Central

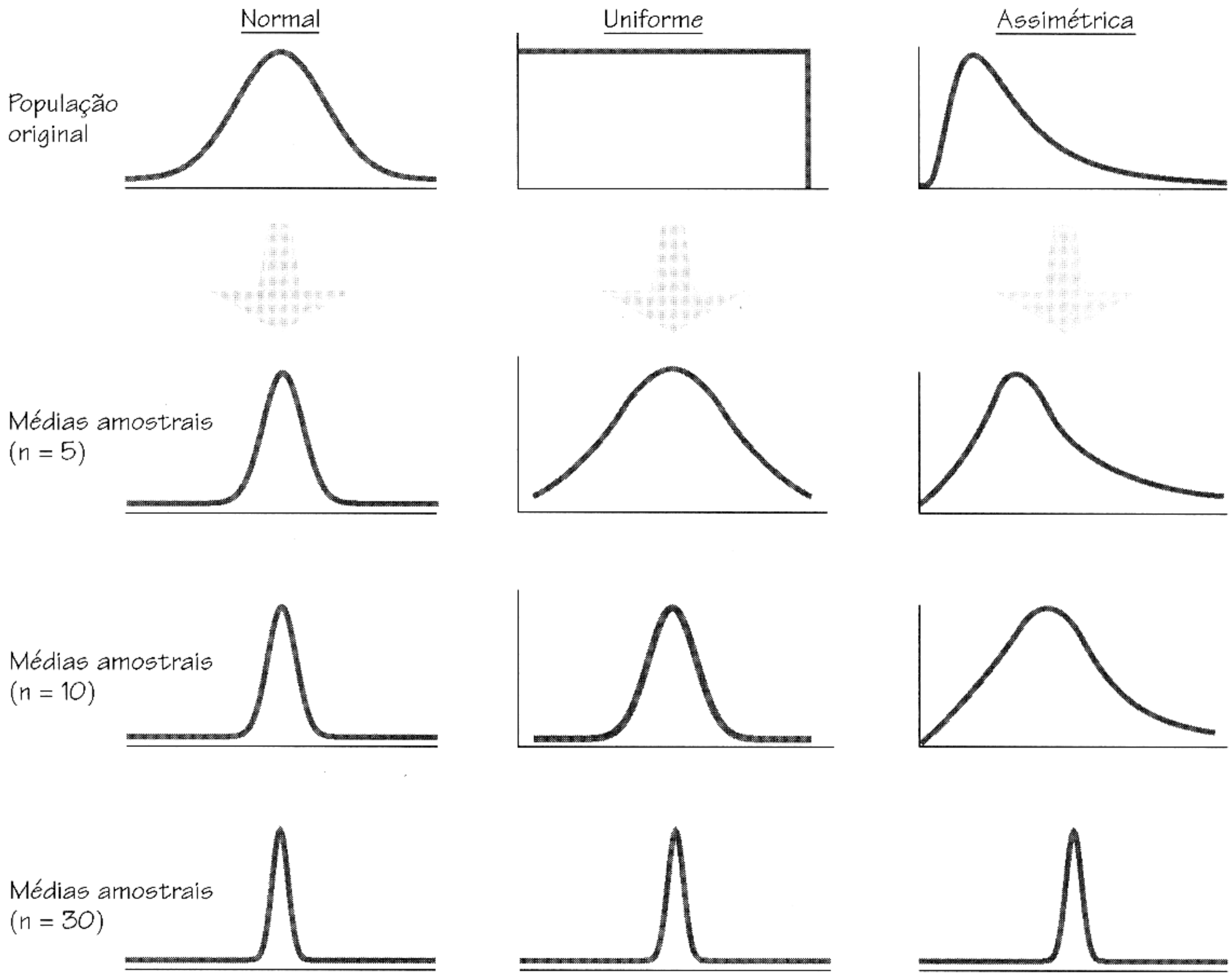


Fig. 5-21 Distribuição normal, uniforme e assimétrica.

Teorema do Limite Central

- 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:

Teorema do Limite Central

- À medida que o tamanho ´n´ da amostra aumenta, a distribuição das médias amostrais <u>tende a uma distribuição normal.</u>

- A <u>média das médias</u> amostrais tenderá à **média populacional:**

$$\mu_{\bar{x}} = \mu = \bar{\bar{x}} = \frac{\sum \bar{x}}{m}$$

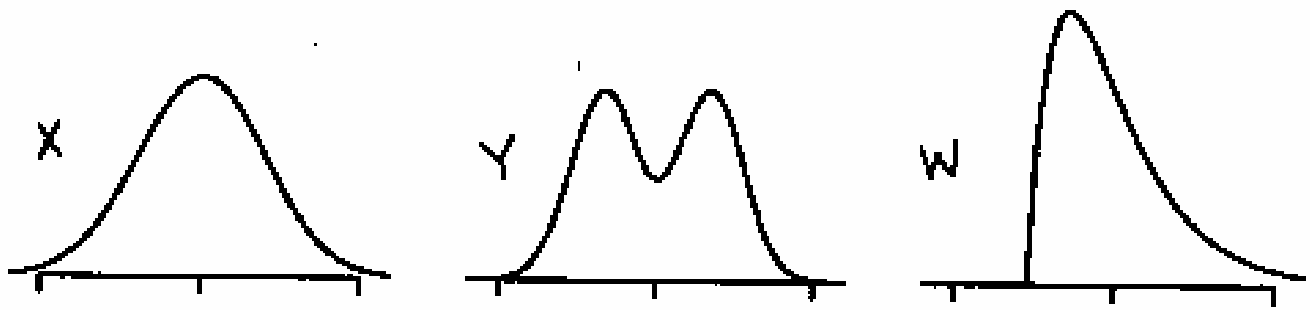
- O desvio padrão das médias amostrais será o **<u>Erro-padrão da média</u>**, dado por:

$$\sigma_{\bar{x}} = \sigma / \sqrt{n}$$

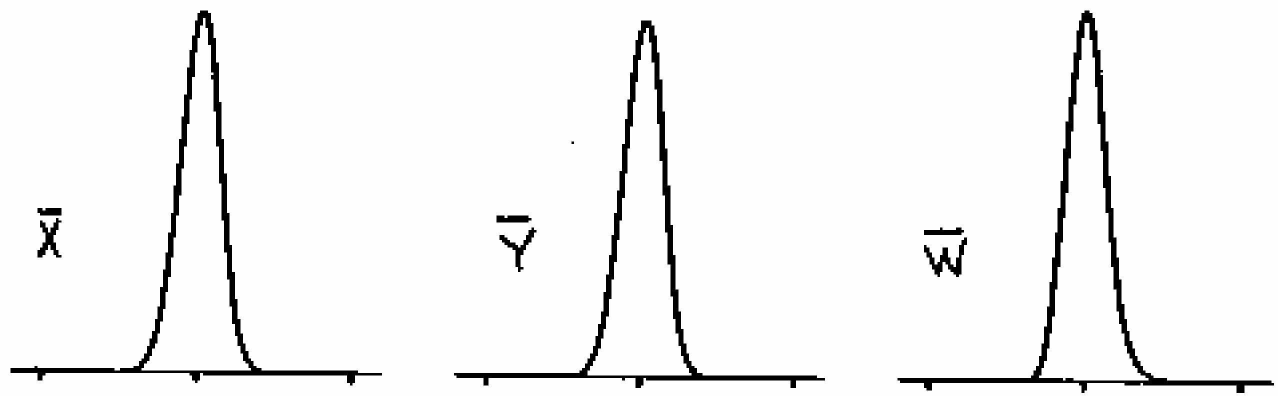
Teorema do Limite Central

- Observações importantes:
  - Quando maior o tamanho das amostras, a distribuição das médias será mais próxima de uma distribuição normal.
  - Regra prática: 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.
- Obs.:
  - 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 <u>estimador pontual</u>.
- 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?
- Associamos, assim, 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

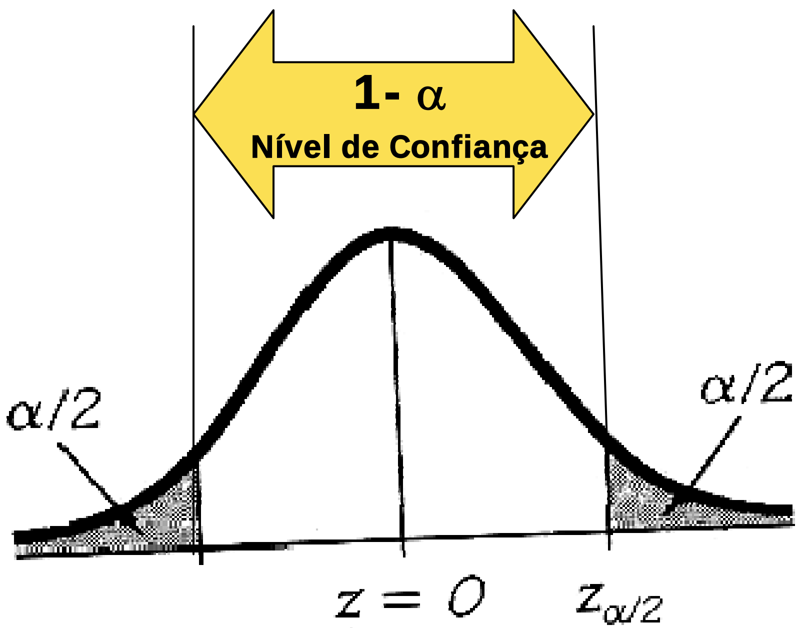
$$\text{Probabilidade}\{c_1 \leq \mu \leq c_2\} = 1 - \alpha$$

- **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> em %.**
- **1- α é o <u>coeficiente de confiança</u>.**

Nível de Confiança (NC)

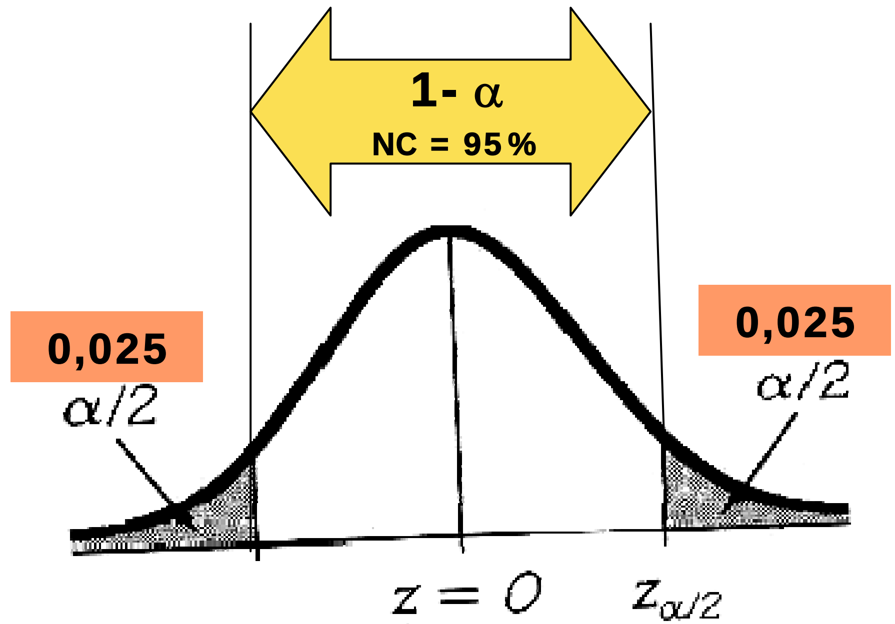
- É a probabilidade **1-α** (comumente expressa percentualmente) do intervalo de confiança conter o valor verdadeiro, o parâmetro populacional

Graças ao Teorema do Limite Central, pode-se usar a Distribuição Normal Padronizada (z) para construir os Intervalos de Confiança (calcular os limites do Intervalo)



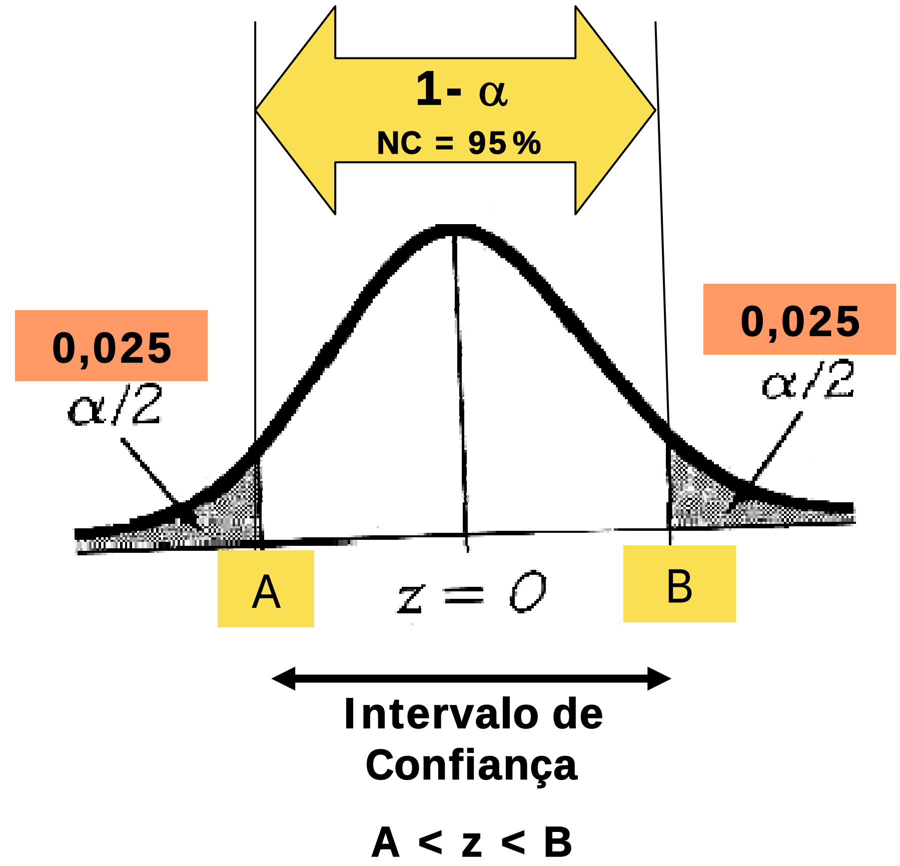
Nível de Confiança (NC)

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



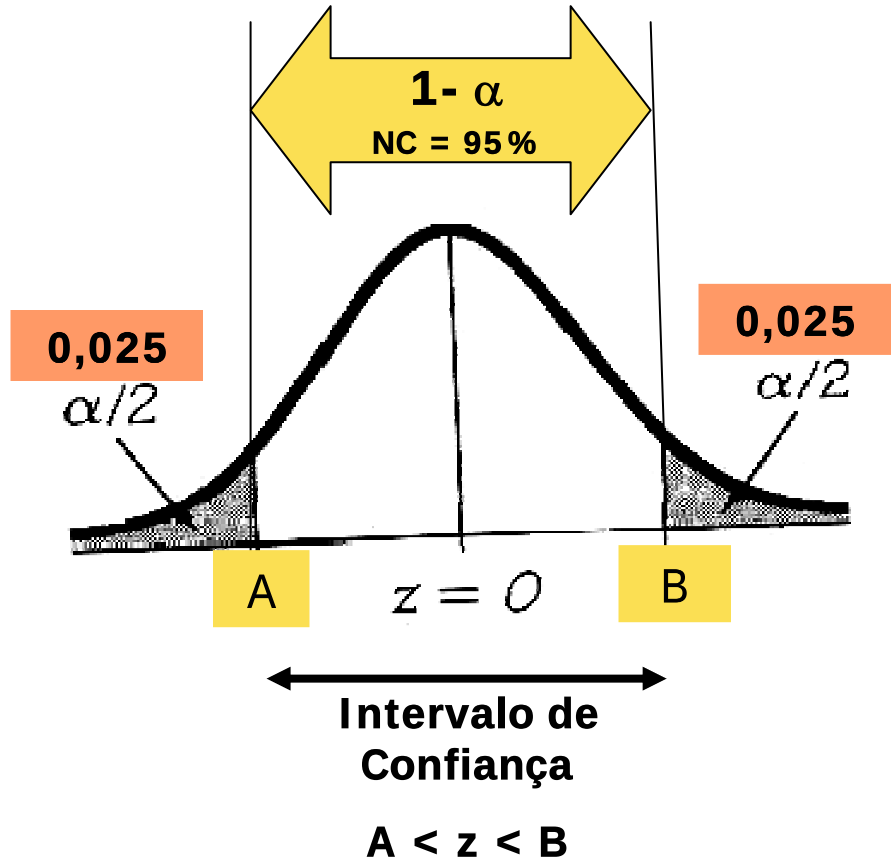
Intervalo de Confiança

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

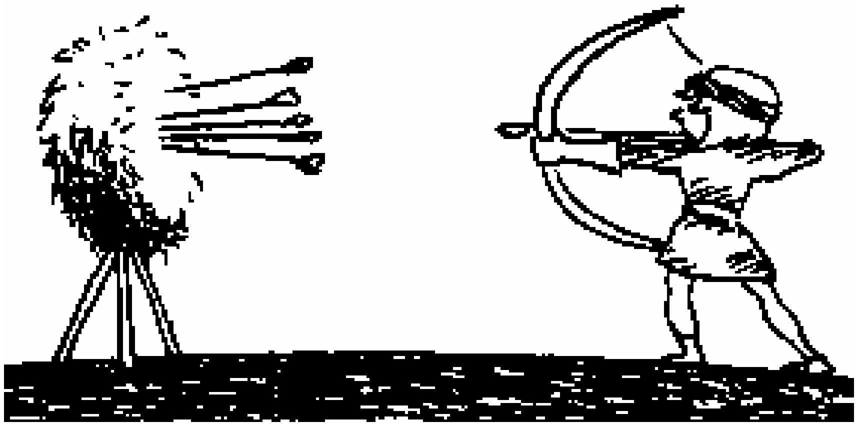


Intervalo de Confiança

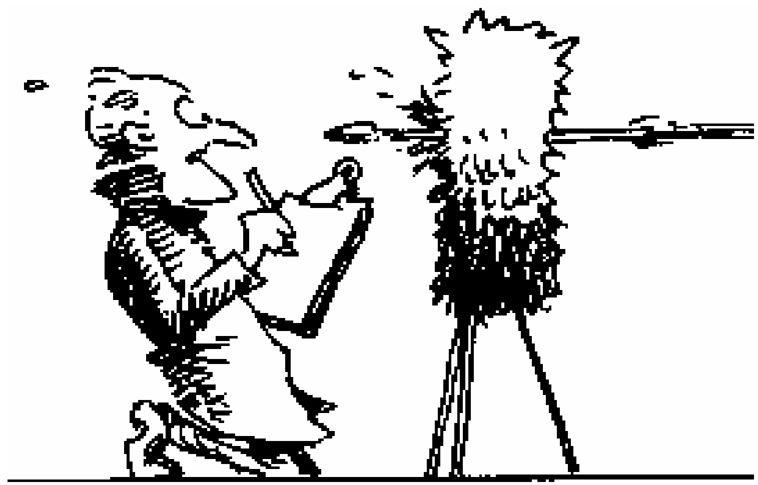
- Conclusão:
  - Se coletarmos várias amostras de 20 alunos e construirmos um intervalo de 95% de confiança para cada uma, a longo prazo, 95% destes intervalos conteriam efetivamente a média da população μ



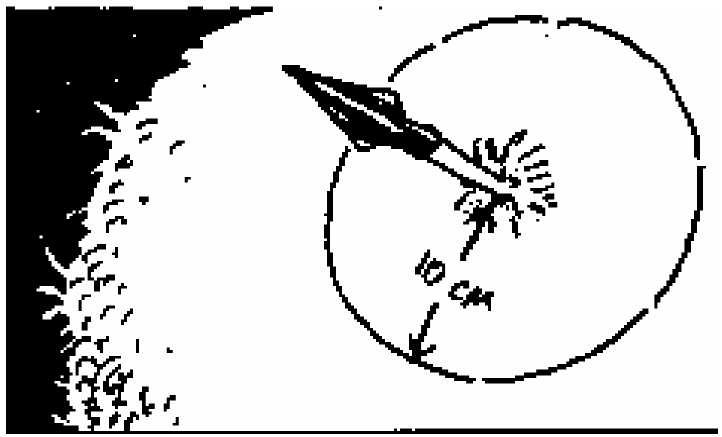
CONSIDERE UMA ARQUEIRA ATIRANDO EM UM ALVO. SUPONHA QUE ELA ACERTA NO CENTRO COM RAIO DE 10 CM 95% DAS VEZES. OU SEJA, ERRA APENAS UMA VEZ A CADA 20 TENTATIVAS.



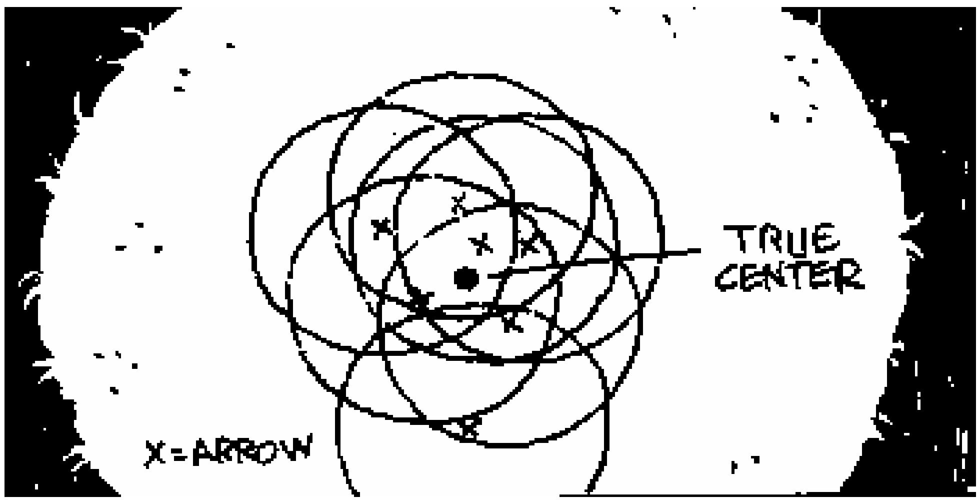
SENTADO ATRÁS DO ALVO ENCONTRA-SE UM BRAVO DETETIVE, QUE NÃO VÊ ONDE ESTÁ O CENTRO. A ARQUEIRA ATIRA A PRIMEIRA FLECHA..



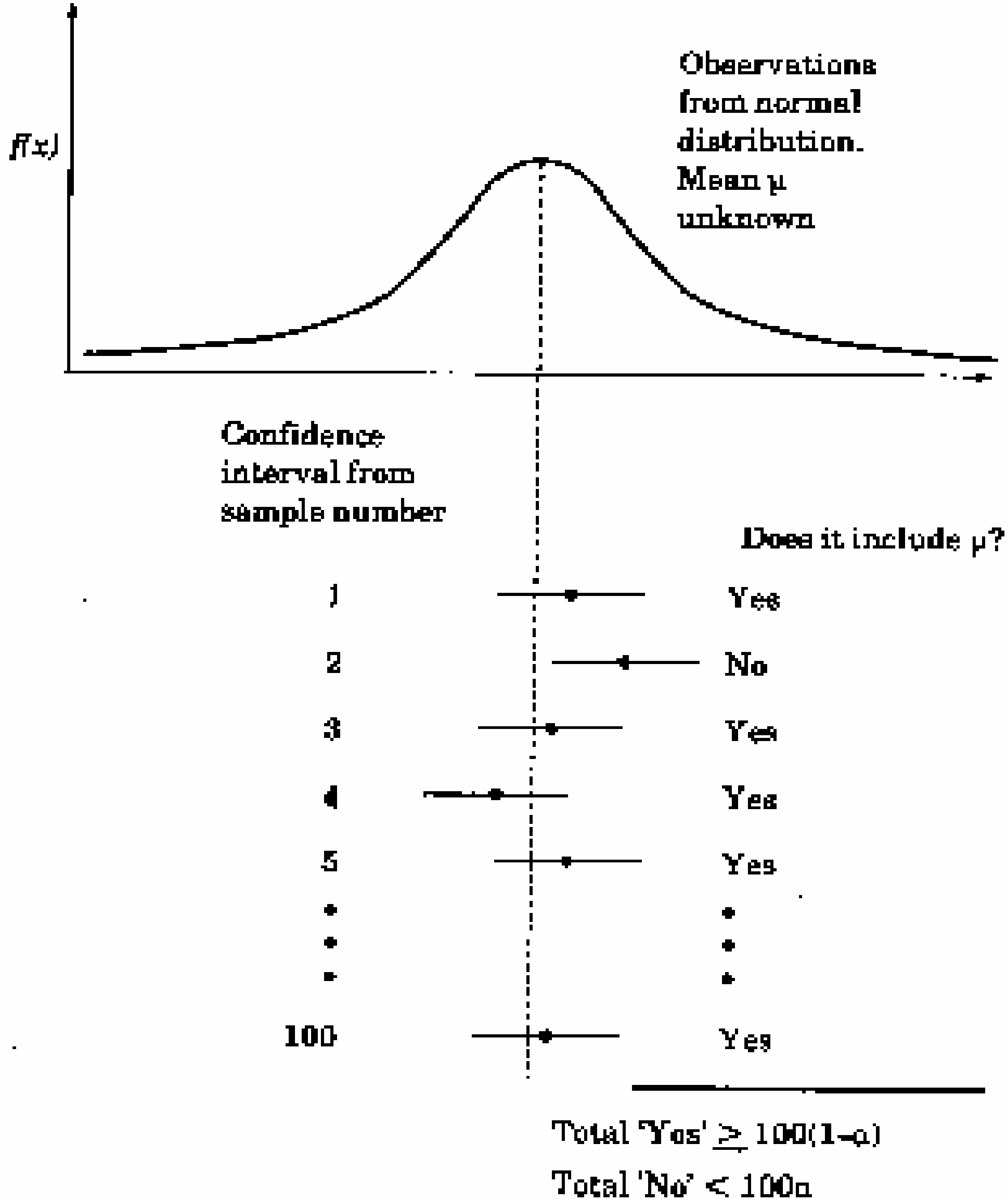
CONHECENDO O NÍVEL DA HABILIDADE DA ARQUEIRA, O DETETIVE DESENHA UM CÍRCULO COM 10 CM DE RAIO AO REDOR DA FLECHA. ELE TEM **95% DE CONFIANÇA** DE QUE O SEU CÍRCULO INCLUI O CENTRO DO ALVO!



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?

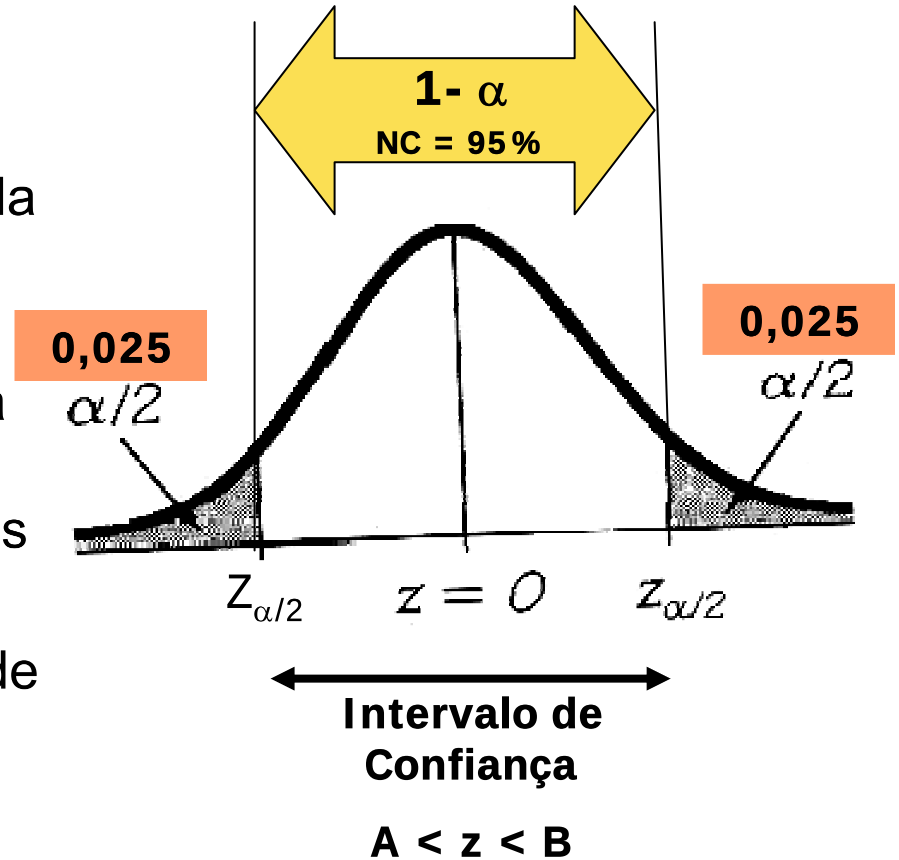
AUMENTANDO O
TAMANHO DO CÍRCULO

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 do Limite Central, 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.
- Por que se usa a Distribuição Normal Padronizada?
  - Pelo Teorema do Limite Central, as médias amostrais distribuem-se normalmente em torno da média das médias. Então, pode-se usar a Normal Padronizada para cálculo das áreas (probabilidades).

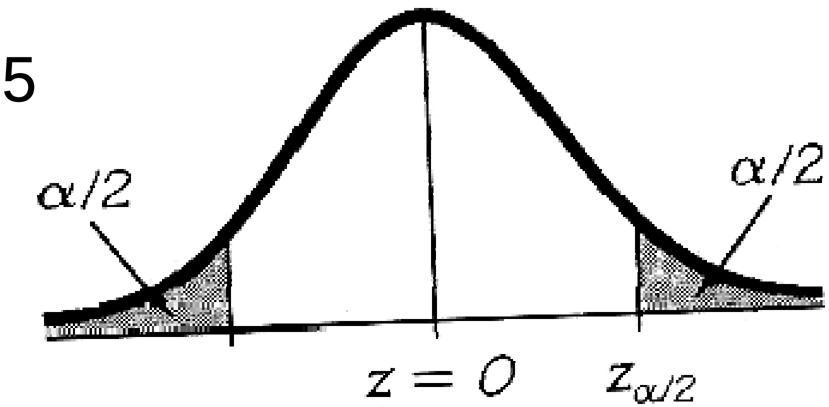
Exercício: Valores $Z_{\alpha/2}$

- 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

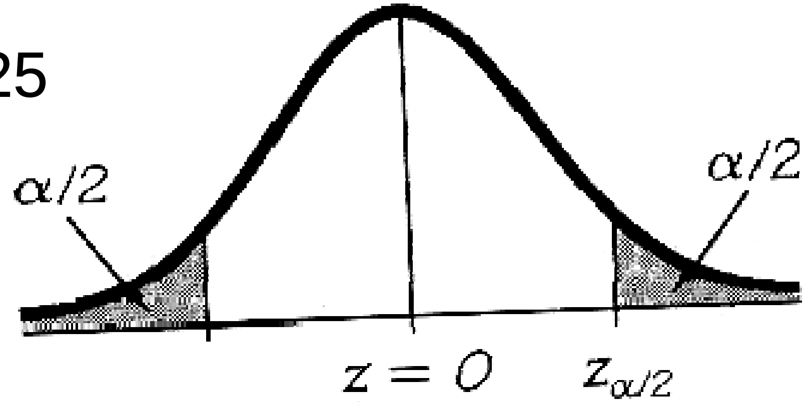
- $\alpha/2$ = 0,05
- Área entre Z=0 e Z=$\alpha/2$ é 0,45
- $Z_{\alpha/2}$ = 1,645



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
- $Z_{\alpha/2} = 1{,}96$

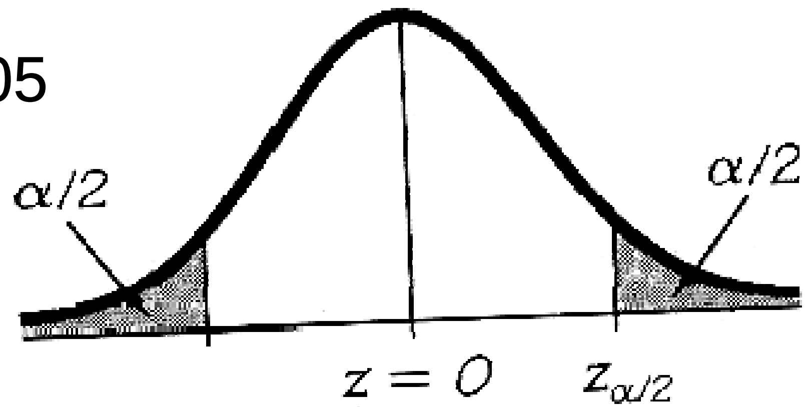
Exercício

- 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

- O processo de produção das unidades de caixa de controle de um tipo de motor foi modificado recentemente. Antes da modificação, os dados históricos indicavam que os diâmetros do orifício dos mancais nas caixas eram distribuídos normalmente com σ=0,100mm. Acredita-se que a modificação no processo não tenha alterado a distribuição ou o desvio padrão, mas o valor do diâmetro médio pode ter mudado.
- Seleciona-se uma amostra de 40 caixas e mede-se o diâmetro do orifício para cada uma, resultando num diâmetro médio de 5,426mm. Calcule um IC para o diâmetro médio real (populacional) do orifício usando um NC de 90%.

Resposta

$$z_{\alpha/2} = \frac{x - \overline{\overline{x}}}{\sigma/\sqrt{n}}$$

$$1{,}645 = \frac{x_s - 5{,}426}{0{,}100/\sqrt{40}} \therefore 0{,}026 = x_s - 5{,}426 \therefore x_s = 5{,}452$$

$$-1{,}645 = \frac{x_i - 5{,}426}{0{,}100/\sqrt{40}} \therefore -0{,}026 = x_i - 5{,}426 \therefore x_i = 5{,}400$$

- O que isto significa?
  - $\mu = 5{,}426 \pm 0{,}026$ ou $5{,}400 < \mu < 5{,}452$
  - Existe 90% de probabilidade do intervalo de 5,400mm a 5,452mm conter a média populacional de diâmetro do orifício do mancal

Exercício

- Na engenharia de produtos, é importante considerar os pesos das pessoas, de modo a evitar sobrecargas (aviões, elevadores) ou falhas (cadeiras que se quebram).
- Dado que a população de homens dos EUA (ano?) tem pesos distribuídos normalmente com média 78,47Kg e desvio-padrão 13,61Kg, determinar a probabilidade de:
  - (a) um homem escolhido aleatoriamente pesar mais de 81,65Kg.
  - (b) em 36 homens escolhidos aleatoriamente, o peso médio ser superior a 81,65Kg.

Solução

- (a) um homem escolhido aleatoriamente pesar mais de 81,65Kg.
- Como trata-se de um *valor individual* proveniente de uma população com distribuição normal, calcular o valor de *z* diretamente:

$$z = \frac{x-\mu}{\sigma} = \frac{81{,}65 - 78{,}47}{13{,}61} = 0{,}2337$$

Da Tabela A-2, a área correspondente a z=0,2337 é **0,0910**. A probabilidade desejada é, pois:

P(z>0,2337) = 0,5 – 0,0910 = **0,4090**

Solução

- (b) em 36 homens escolhidos aleatoriamente, o peso médio ser superior a 81,65Kg.
- Como estamos lidando com a *média para um grupo* de 36 valores, usamos o *Teorema do Limite Central* (cada valor individual seria uma amostra?)

$$\mu_{\bar{x}} = \mu = 78{,}47$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{13{,}61}{\sqrt{36}} = 2{,}2683$$

O escore *z* de interesse é agora calculado:

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{81{,}65 - 78{,}47}{13{,}61 / \sqrt{36}} = 1{,}4019 \Rightarrow P(z > 1{,}4019) = 0{,}5 - 0{,}4192 = 0{,}0808$$

Comentários

- Há uma probabilidade de 0,4090 de um homem pesar mais que 81,65Kg, mas a probabilidade de 36 homens terem peso médio superior a 81,65Kg é de apenas 0,0808!
- É muito mais fácil um único indivíduo afastar-se da média, do que um grupo de 36 indivíduos.