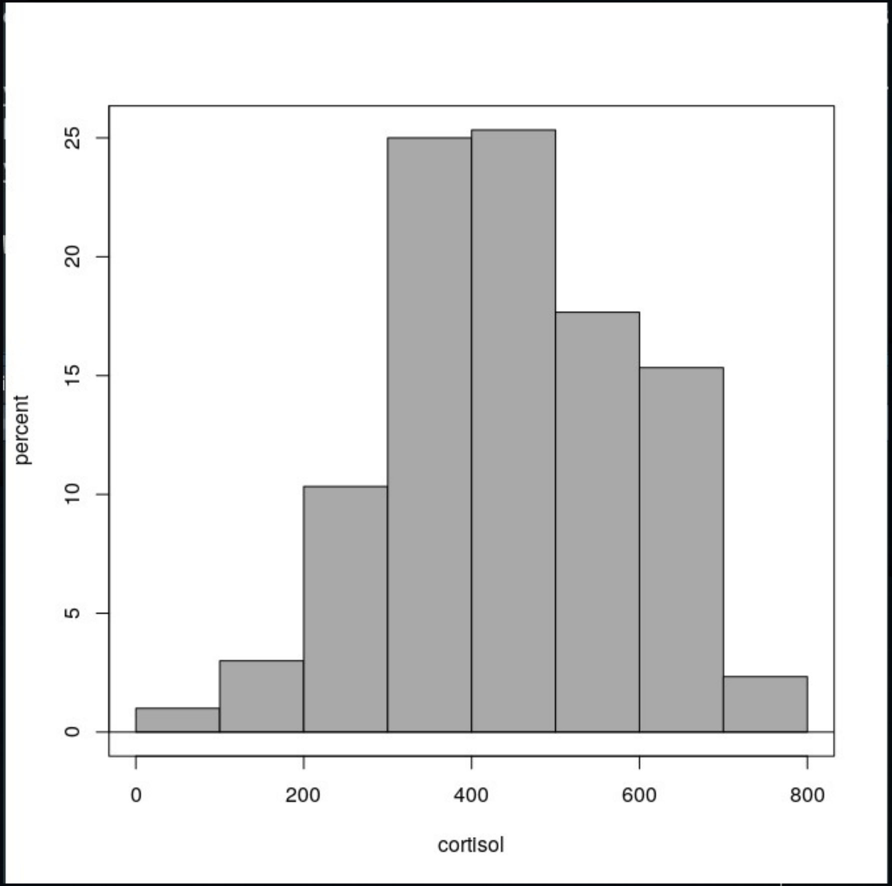
Distribuições e Teorema do Limite Central

- Distribuição empírica de frequência
- Distribuição Binomial
- Distribuição Normal
- Teorema do Limite Central
- Estimativa de média populacional com uma amostra

Distribuição empírica de frequência



Distribuições de probabilidade

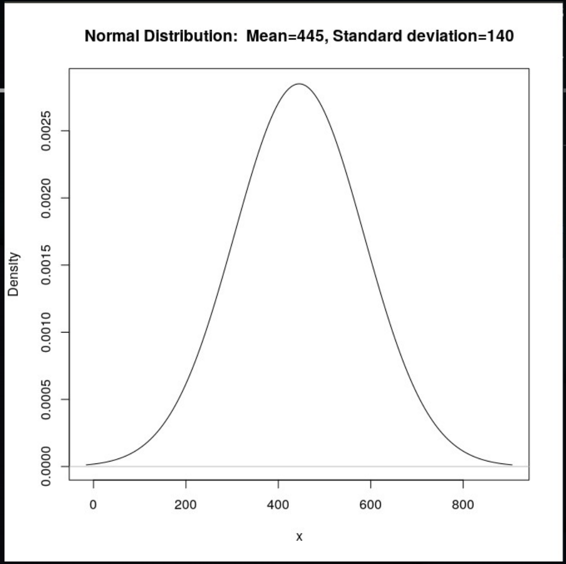
- Princípio teórico:

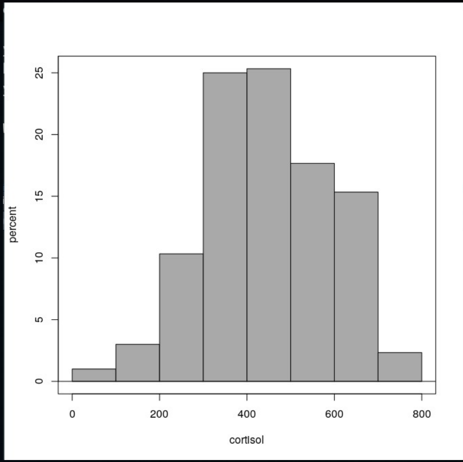
  “Existe uma função que governa a probabilidade de obtermos determinados valores na observação de uma grandeza” (Helene e Vanin)

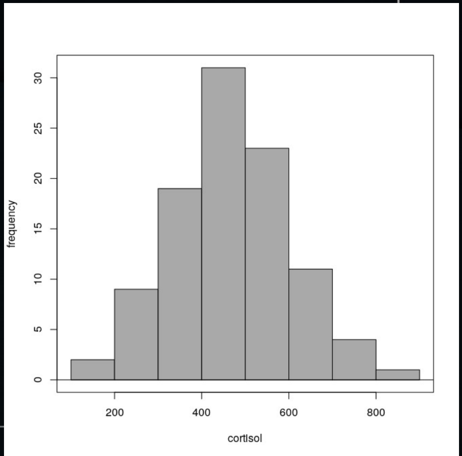
$$\Downarrow$$

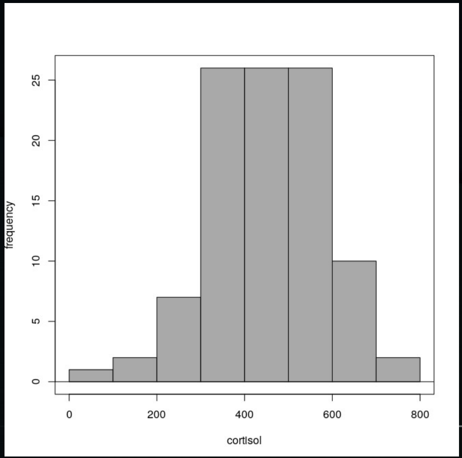
Função (Densidade) de Probabilidade (fdp)

“Podemos entender uma distribuição de probabildiade como um equivalente teórico de uma distribuição empírica de frequências” (Petrie e Watson)


Normal Distribution: Mean=445, Standard deviation=140
Density
0.0000
0.0005
0.0010
0.0015
0.0020
0.0025
0
200
400
600
800
x


percent
0
5
10
15
20
25
0
200
400
600
800
cortisol

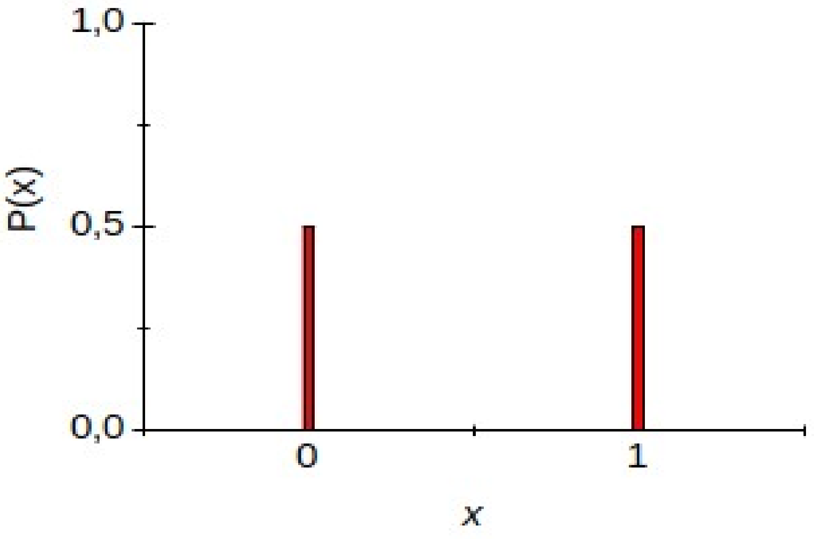

frequency
0
5
10
15
20
25
30
200
400
600
800
cortisol


frequency
0
5
10
15
20
25
0
200
400
600
800
cortisol

Distribuição binomial

- Distribuição de valores discretos mais conhecida
- O resultado de *n* observações de uma variável dicotômica

Distribuição binomial



Evento	x	P(x)
F	0	q=1/2
M	1	p=1/2

x: variável aleatória que representa o número de bezerros nascidos do sexo masculino (M) em n nascidos

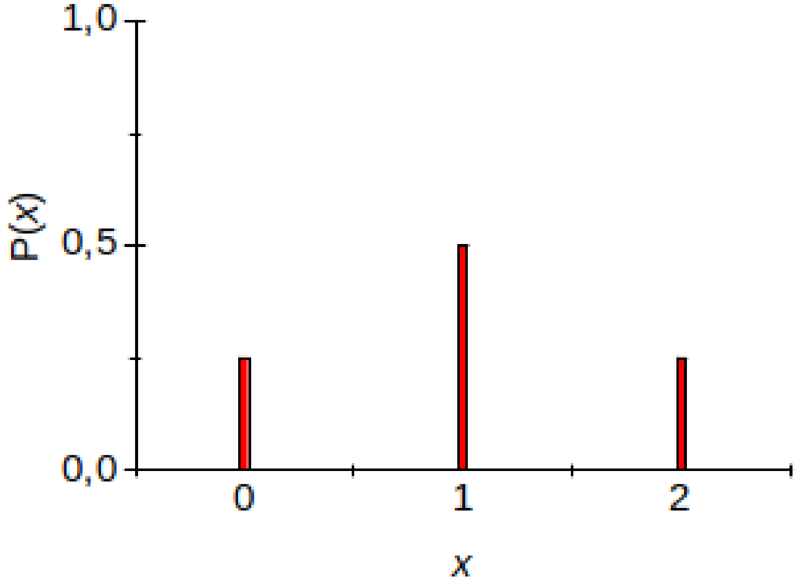
p: probabilidade de nascer um bezerro macho

$q=1-p$: probabilidade de nascer fêmea

n=1

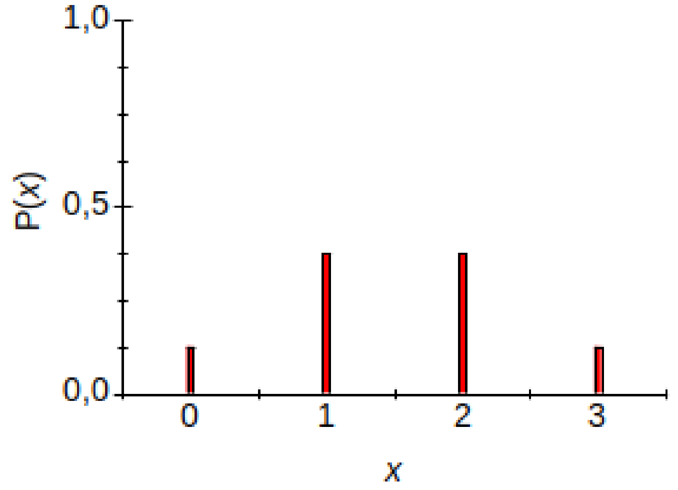
p=1/2

Distribuição binomial



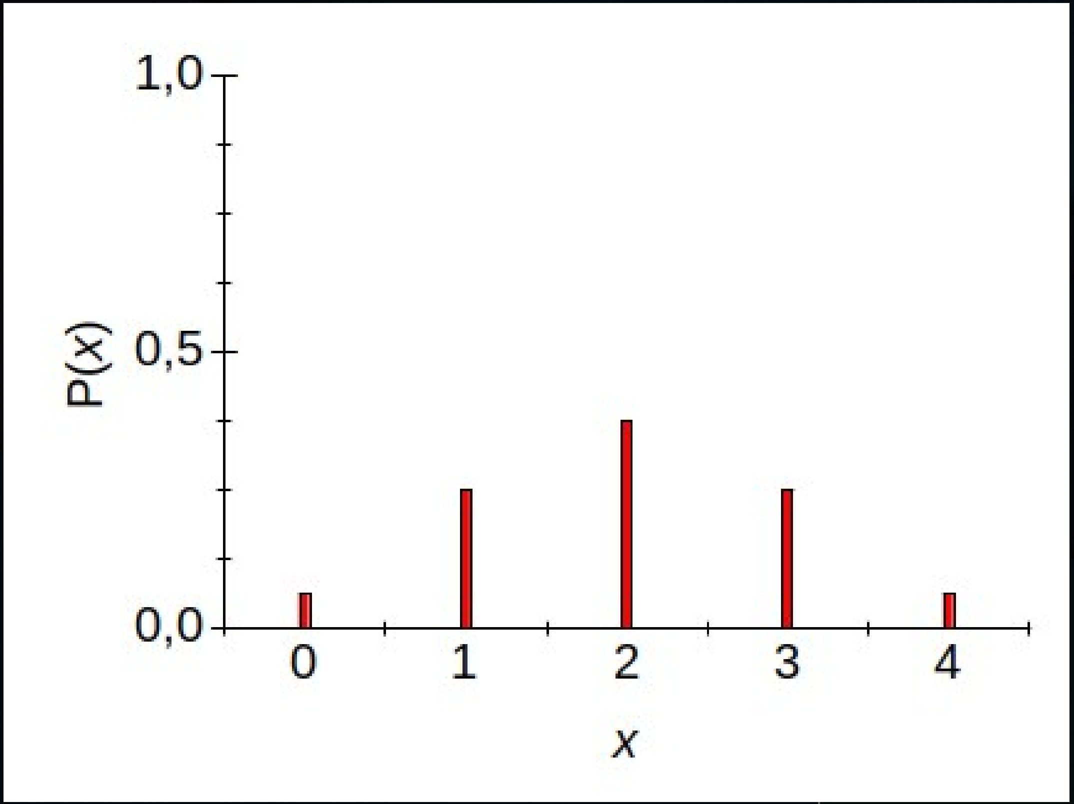
Evento	*x*	P(*x*)	
FF	0	q.q=1/4	
FM	1	q.p	2q.p=1/2
MF	1	p.q	
MM	2	p.p=1/4	

Distribuição binomial



Evento	x	P(x)	
FFF	0	q.q.q = 1/8	
FFM	1	q.q.p	$3q^2p = 3/8$
FMF	1	q.p.q	
MFF	1	p.q.q	
FMM	2	q.p.p	$3qp^2 = 3/8$
MFM	2	p.q.p	
MMF	2	p.p.q	
MMM	3	p.p.p=1/8	

Distribuição binomial



Distribuição binomial

$$P(x) = \binom{n}{x} p^x q^{n-x}$$

n: número de observações

x: número de eventos de um certo tipo (“sucesso”)

p: probabilidade de ocorrência do evento que nos interessa

média: $\mu = n\,p$

variância: $\sigma^2 = n\,p\,q$

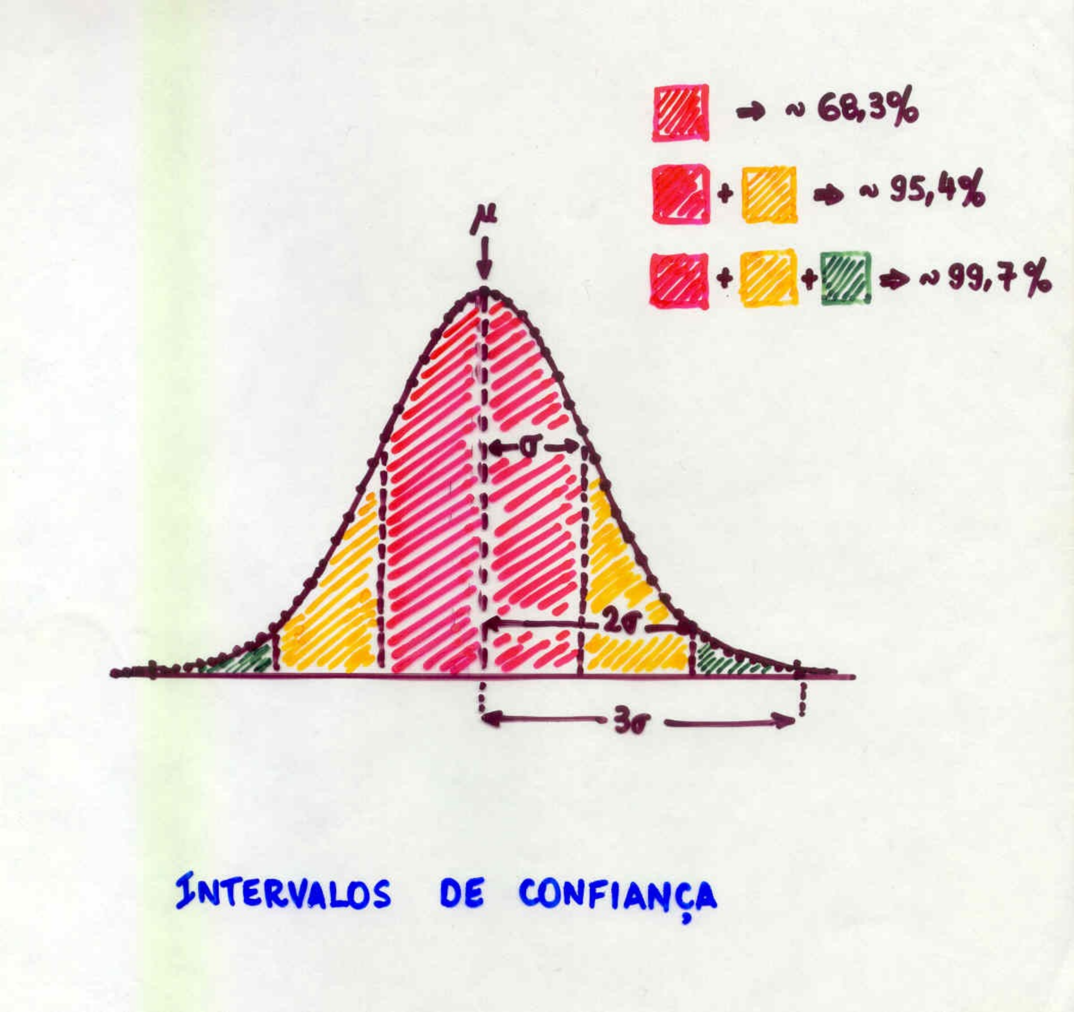
Distribuição Normal

ou Gaussiana

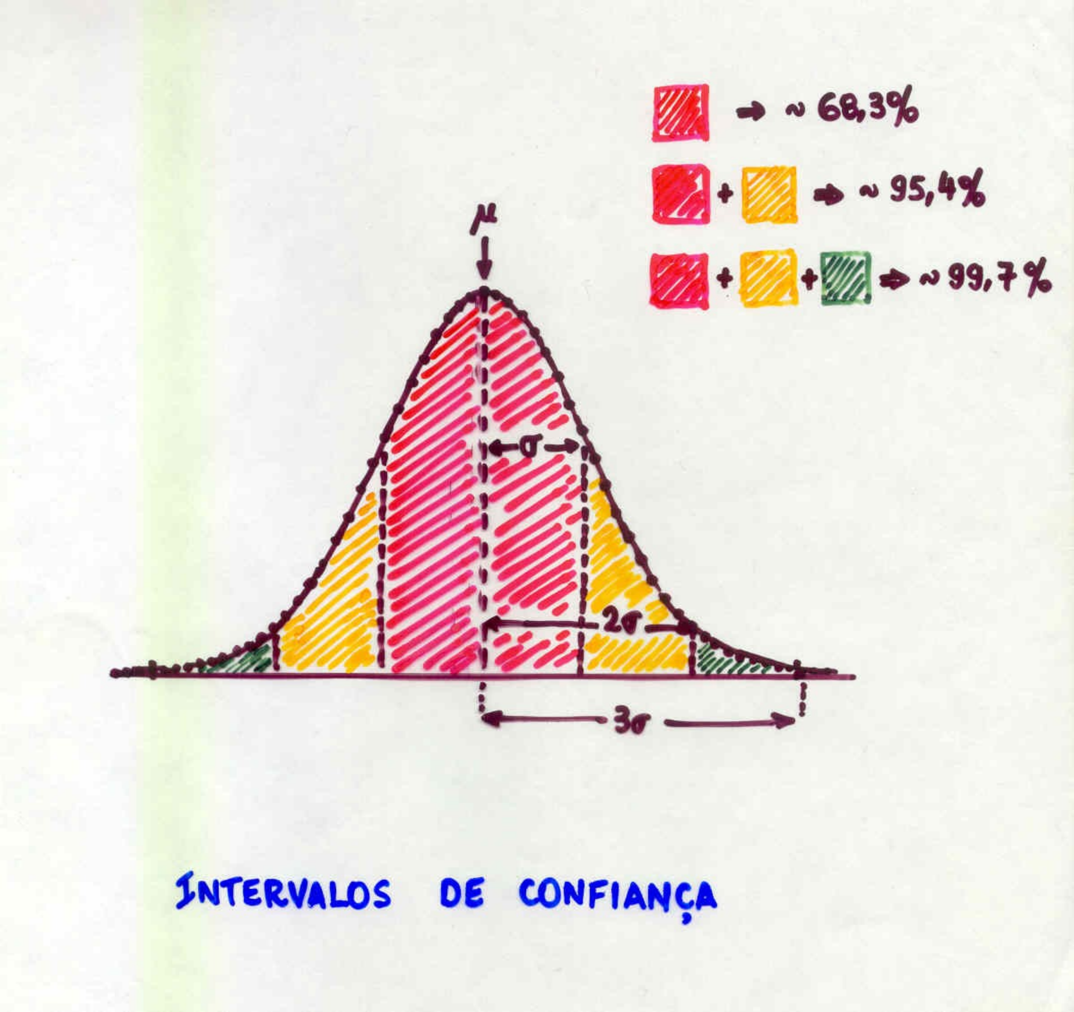
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

- Descrita por 2 parâmetros
  - Média, desvio padrão
- Unimodal
- Simétrica em torno da média
- Forma de sino
- Média = mediana = moda

Distribuição Normal



Distribuição Normal



$$P(\mu - 1{,}96\sigma < x < \mu + 1{,}96\sigma) = 95\%$$

$$z = 1{,}96$$

Teorema do Limite Central

- “Se tomarmos amostras grandes de uma população, as médias amostrais terão distribuição Normal mesmo que os dados originais não tenham distribuição Normal.”

- Simulações:
- http://onlinestatbook.com/stat_sim/sampling_dist/index.html

Teorema do Limite Central

Máquina de Galton

- Vídeo Máquina de Galton: https://youtu.be/sPHDZLd4vmU

- Média da distribuição de médias

$$\bar{\bar{x}} = \bar{x}_m = \mu$$

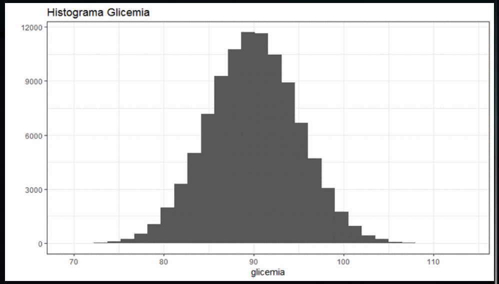
- Desvio padrão das médias

$$s_m = \frac{\sigma}{\sqrt{n}}$$

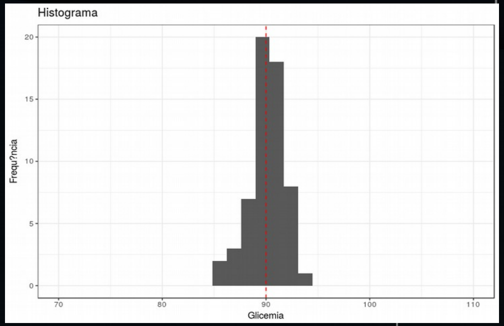
μ = média da população
σ = desvio padrão da população
n = tamanho da amostra

Estimativa de média populacional

População



Médias amostrais

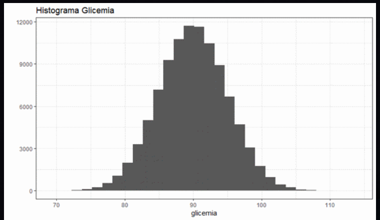


É possível calcular as médias amostrais prováveis que podem ser obtidas a partir de uma população com média populacional conhecida

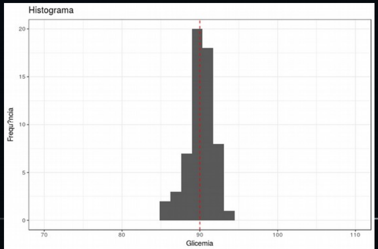
Estimativa de média populacional

De maneira análoga, é possível calcular as médias populacionais prováveis que podem ter gerado uma amostra com dada média amostral

População



Médias amostrais



Intervalo de confiança

Médias Populacionais mais prováveis

Média amostral

Da mesma maneira que podemos calcular as médias amostrais possíveis de uma população,

Podemos calcular as populações possíveis que geraram uma média amostral

Da normal: $P(\mu - 1{,}96\ \sigma < x < \mu + 1{,}96\ \sigma) = 95\,\%$

No TLC: $P(\mu - 1{,}96 * \sigma_m < \overline{x} < \mu + 1{,}96 * \sigma_m) = 95\,\%$

$$\Downarrow$$

$P(\overline{x} - 1{,}96 * \sigma_m < \mu < \overline{x} + 1{,}96 * \sigma_m) = 95\,\%$

Tamanho do erro: $\varepsilon = 1{,}96 * s_m$

Ou seja, tamanho do intervalo de confiança (IC), do erro, depende do erro-padrão.

$$s_m = \frac{\sigma}{\sqrt{n}}$$

Portanto:
Quanto maior o n (tamanho da amostra), menor o IC
Quanto maior o desvio-padrão da população, maior o IC

Tamanho da amostra

$$\varepsilon = 1{,}96 * s_m$$

$$\varepsilon = 1{,}96 * \frac{\sigma}{\sqrt{n}} \quad \Rightarrow \quad n = 1{,}96^2 * \frac{\sigma^2}{\varepsilon^2}$$

Obrigado