Parte 03

Estimação de parâmetros

Prof. Dr. Renato de Oliveira Moraes

Estimador e Estimativa

- θ = parâmetro a ser estimado
- T = um estimador de θ
- t = uma dada estimativa

Propriedades dos estimadores

Justeza (não tendenciosidade)

$$\mu(T) = \theta$$

Consistência

$$\lim_{n \to \infty} P\left(\left|T - \theta\right| \geq \varepsilon\right) = 0$$

ou

$$\lim_{n \to \infty} \sigma^2(T) = 0$$

Eficiência

$$\mu\left[(T_1 - \theta)^2\right] < \mu\left[(T_2 - \theta)^2\right] \Rightarrow T_1 \text{ é mais eficiente que } T_2$$

Estimador para a variância populacional σ^2

Estimador θ para a variância populacional σ²

$$\theta = \frac{\boxed{\sum\left(x_i - \bar{x}\right)^2}}{N} = \frac{\sum\left(x_i - \mu + \mu - \bar{x}\right)^2}{N}$$

$$\sum\left(x_i - \mu + \mu - \bar{x}\right)^2 = \sum\left((x_i - \mu)^2 + 2(x_i - \mu)\left(\mu - \bar{x}\right) + \left(\mu - \bar{x}\right)^2\right)$$

$$\sum(x_i - \mu)^2 + \sum 2(x_i - \mu)\left(\mu - \bar{x}\right) + \sum\left(\mu - \bar{x}\right)^2 =$$

$$\sum(x_i - \mu)^2 - 2\left(\mu - \bar{x}\right)\sum(\mu - x_i) + \sum\left(\mu - \bar{x}\right)^2 =$$

$$\sum(x_i - \mu)^2 - 2\left(\mu - \bar{x}\right)\left(N\mu - N\bar{x}\right) + \left(\mu - \bar{x}\right)^2\sum 1 =$$

$$\sum(x_i - \mu)^2 - 2N\left(\mu - \bar{x}\right)^2 + N\left(\mu - \bar{x}\right)^2 = \boxed{N\sigma^2 - N\left(\mu - \bar{x}\right)^2}$$

Estimador θ para a variância populacional σ^2

$$\theta = \frac{\sum \left(x_i - \bar{x}\right)^2}{N} = \frac{N\sigma^2 - N\left(\mu - \bar{x}\right)^2}{N}$$

$$E[\theta] = E\left[\frac{N\sigma^2 - N\left(\mu - \bar{x}\right)^2}{N}\right] = \frac{E\left[N\sigma^2 - N\left(\mu - \bar{x}\right)^2\right]}{N}$$

$$E[\theta] = \frac{E\left[N\sigma^2\right] - NE\left[\left(\mu - \bar{x}\right)^2\right]}{N} = \frac{N\sigma^2 - N\dfrac{\sigma^2}{N}}{N} = \frac{N\sigma^2 - \sigma^2}{N}$$

$$E[\theta] = \sigma^2 \frac{(N-1)}{N}$$

Estimador de máxima verossimilhança

Qual o valor do parâmetro a ser estimado que maximiza a probabilidade do estimador assumir o valor observado na amostra?

$$L(\theta) = \max P(x_1, x_2, x_3 \ldots x_n \mid \theta)$$

Estimadores:

- Consistentes
- Assintoticamente eficientes
- Distribuição assintoticamente Normal

Exemplo

Uma caixa com 10 bolas, das quais S são pretas e 10 – S são brancas. Uma amostra de 4 bolas com extraídas com reposição foi colhida e observou-se que três bolas eram brancas e apenas uma era. O número bolas pretas na amostra tem uma distribuição Binomial. A função de verossimilhança fica:

$$L(S)=\binom{4}{1}\frac{S}{10}\left(\frac{10-S}{10}\right)^3=4\frac{S}{10}\left(\frac{10-S}{10}\right)^3$$

Nestas condições o máximo valor da função de verossimilhança ocorre quando S = 3.

S	L(S)
0	0,00%
1	29,16%
2	40,96%
3	41,16%
4	34,56%
5	25,00%
6	15,36%
7	7,56%
8	2,56%
9	0,36%
10	0,00%

Exercício

- Uma única observação é efetuada de uma variável aleatória discreta X com função de probabilidade $f(x|\theta)$, onde $\theta \in \{1,2,3\}$. Encontre o Estimador de Máxima verossimilhança de θ.

X	$f(x\|\theta=1)$	$f(x\|\theta=2)$	$f(x\|\theta=3)$
0	1/3	¼	0
1	1/3	¼	0
2	0	¼	¼
3	1/6	¼	½
4	1/6	0	¼

X	Estimativa (valor de θ)
0	1
1	1
2	2 ou 3
3	3
4	4

Estimador de máxima verossimilhança

- Amostra com N elementos: $x_1, x_2, x_3 \dots x_n$
- Calcula-se a probabilidade de se obter a amostra observada em função do valor do parâmetro que se deseja estimar – $f(x_i, \theta)$
- $P(amostra) = f(x_1) \times f(x_2) \times \cdots \times f(x_n)$
- $P(amostra) = L(\theta) = f(x_1) \times f(x_2) \times \cdots \times f(x_n)$
- Aplica-se a função logarítmica a $L(\theta)$ e procura-se o valor de θ que maximiza o valor da função $ln\big(L(\theta)\big)$
- Deriva e iguala a zero: $\frac{\partial ln\big(L(\theta)\big)}{\partial \theta} = 0$

Distribuição de Bernoulli

Qual é o estimador de p? (probabilidade de sucesso)

$$f(x) = p^x(1-p)^{1-x}; x = \begin{cases} 1, P(X) = p \\ 0, P(X) = 1-p \end{cases}$$

$$L(p) = f(x_1) \times f(x_2) \times \cdots \times f(x_N) = \prod_{i=1}^{N} f(x_i)$$

$$L(p) = p^{\sum_{i=1}^{N} x_i}(1-p)^{\sum_{i=1}^{N} 1-x_i} = p^{\sum_{i=1}^{N} x_i}(1-p)^{N-\sum_{i=1}^{N} x_i}$$

$$ln\big(L(p)\big) = ln\left(p^{\sum_{i=1}^{N} x_i}(1-p)^{N-\sum_{i=1}^{N} x_i}\right)$$

$$ln\big(L(p)\big) = ln\left(p^{\sum_{i=1}^{N} x_i}\right) + ln\left((1-p)^{N-\sum_{i=1}^{N} x_i}\right)$$

$$ln\big(L(p)\big) = \left(\sum_{i=1}^{N} x_i\right) ln(p) + \left(N - \sum_{i=1}^{N} x_i\right) ln(1-p)$$

- $ln\left(L(p)\right) = \left(\sum_{i=1}^{N} x_i\right) ln(p) + \left(N - \sum_{i=1}^{N} x_i\right) ln(1-p)$

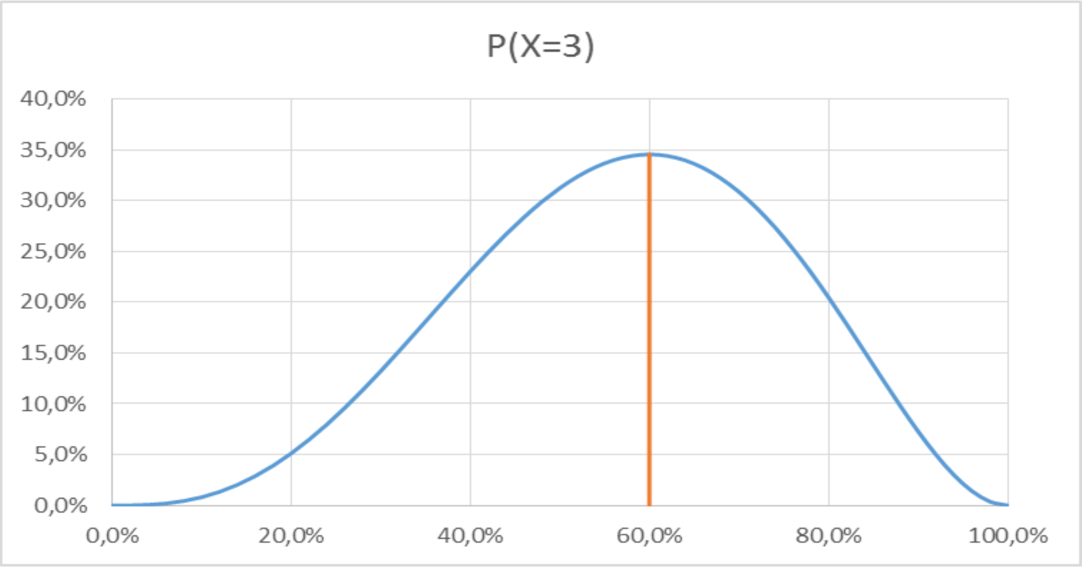
- $\frac{\partial ln\left(L(p)\right)}{\partial p} = 0 \Longrightarrow$

- $\frac{\sum_{i=1}^{N} x_i}{p} - \frac{N - \sum_{i=1}^{N} x_i}{1-p} = 0 \Longrightarrow \frac{\sum_{i=1}^{N} x_i}{p} = \frac{N - \sum_{i=1}^{N} x_i}{1-p} \Longrightarrow$
- $(1-p) \times \left(\sum_{i=1}^{N} x_i\right) = p \times \left(N - \sum_{i=1}^{N} x_i\right) \Longrightarrow$
- $p \times \left(N - \sum_{i=1}^{N} x_i + \sum_{i=1}^{N} x_i\right) = \sum_{i=1}^{N} x_i \Longrightarrow$

- $\hat{p} = \frac{\sum_{i=1}^{N} x_i}{N}$

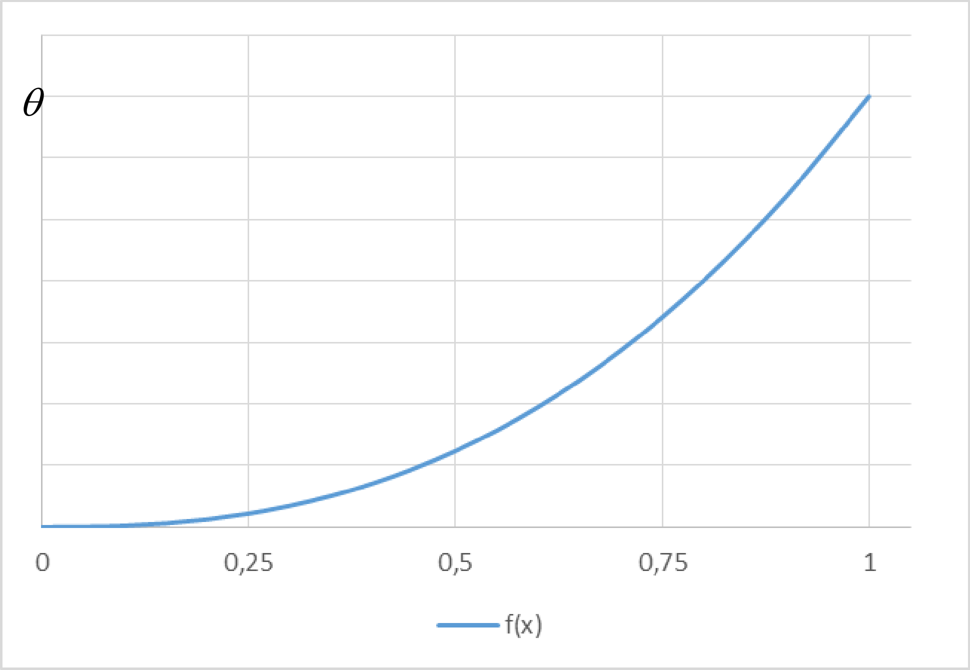
Exemplo 1

- Amostra de 5 elementos {1; 0; 1; 1; 0}
- {sucesso; fracasso; sucesso; sucesso; fracasso}
- $\hat{p} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{3}{5}$



Distribuição Contínua

$$f(x) = \theta x^{(\theta-1)}$$

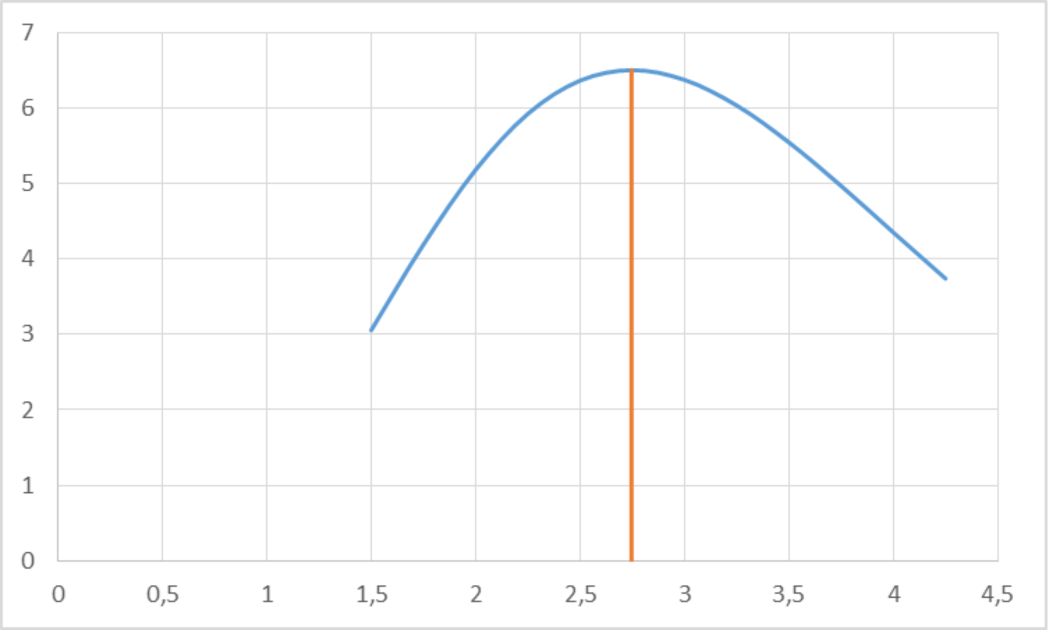


Distribuição Contínua

- $f(x) = \theta x^{(\theta-1)}$
- Amostra de N elementos
- $L(\theta) = f(x_1) \times f(x_2) \times \cdots \times f(x_N) = \left(\theta x_1^{(\theta-1)}\right) \times \left(\theta x_2^{(\theta-1)}\right) \times \cdots \times \left(\theta x_n^{(\theta-1)}\right) = \theta^N \times (x_1 \times x_2 \times \cdots \times x_N)^{(\theta-1)}$
- $ln\big(L(\theta)\big) = Nln(\theta) + (\theta - 1) \times ln(x_1 \times x_2 \times \cdots \times x_N)$
- $ln\big(L(\theta)\big) = Nln(\theta) + (\theta - 1) \times ln\left(\prod_{i=1}^{N} x_i\right)$
- $\frac{\partial ln\left(L(\theta)\right)}{\partial \theta} = 0 \Longrightarrow \frac{N}{\theta} + ln\left(\prod_{i=1}^{N} x_i\right) \Longrightarrow \hat{\theta} = \frac{-N}{ln\left(\prod_{i=1}^{N} x_i\right)}$

Exemplo

- Amostra de 5 elementos {0,37; 0,88; 0,94; 0,67; 0,79}
- $\hat{\theta} = \frac{-N}{ln\left(\prod_{i=1}^{N} x_i\right)}$
- $\prod_{i=1}^{5} x_i = 0{,}162$
- $ln\left(\prod_{i=1}^{N} x_i\right) = -1{,}82$
- $\hat{\theta} = \frac{-5}{-1{,}82} = 2{,}75$

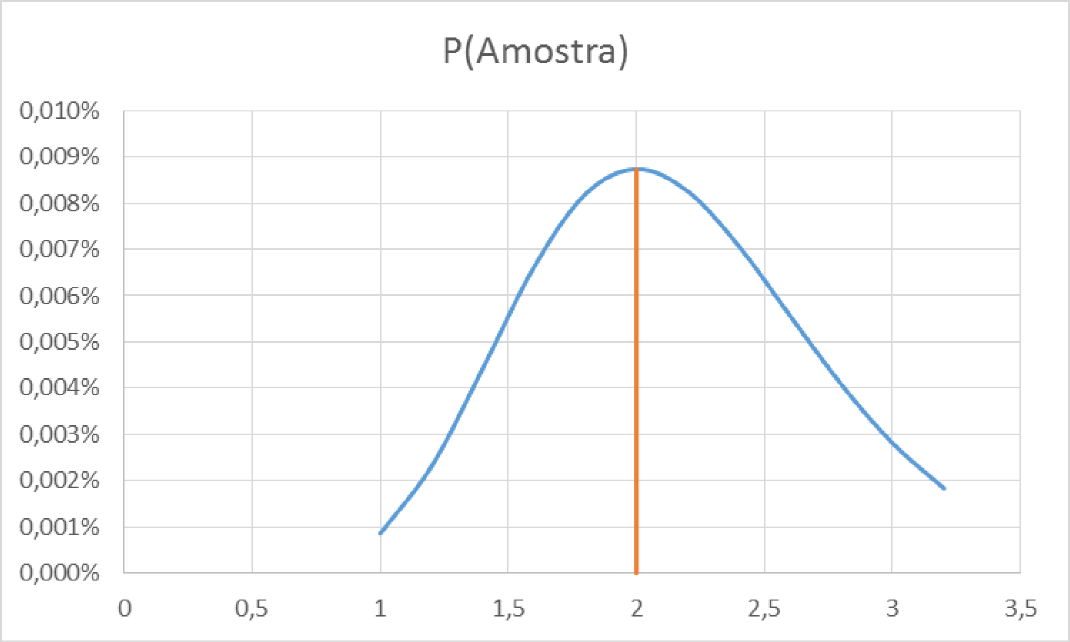


Distribuição de Poisson

- Amostra de N elementos, deseja-se estimar o valor de μ (λt)
- $f(x) = \frac{\mu^x e^{-\mu}}{x!} \Longrightarrow L(\mu) = \prod_{i=1}^{N}\left(\frac{\mu^{x_i} e^{-\mu}}{x_i!}\right) = \frac{\mu^{\sum_{i=1}^{N} x_i}}{\prod_{i=1}^{N}(x_i!)} e^{-N\mu}$
- $ln\big(L(\mu)\big) = \left(\sum_{i=1}^{N} x_i\right) ln(\mu) - \left(\sum_{i=1}^{N}\big(ln(x_i)\big)\right) - (N\mu)$
- $\frac{\partial ln\big(L(\mu)\big)}{\partial \mu} = 0 \Longrightarrow \left(\sum_{i=1}^{N} x_i\right)\frac{1}{\mu} - 0 - N = 0$
- $\hat{\mu} = \frac{\sum_{i=1}^{N} x_i}{N}$

Exemplo com distribuição de Poisson

Amostra com 6 elementos {4; 3; 1; 2; 1; 1} → $\bar{x}$ = 2

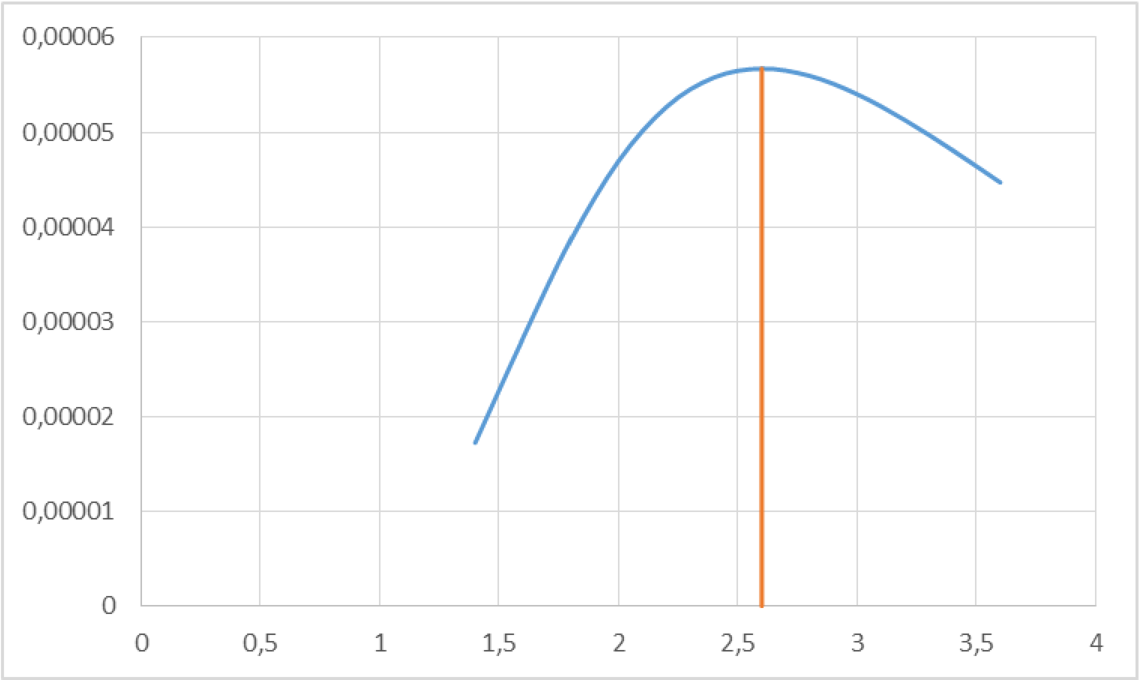


Distribuição Exponencial

- Amostra com N elementos, deseja-se estimar a taxa de ocorrência $\lambda = {}^{1}/_{E[X]}$
- $f(x) = \lambda e^{-\lambda x} \Longrightarrow L(\lambda) = \prod_{i=1}^{N} f(x_i) = \prod_{i=1}^{N} \left(\lambda e^{-\lambda x_i}\right) = \lambda^N e^{-\lambda \sum_{i=1}^{N} x_i}$
- $ln\big(L(\lambda)\big) = ln(\lambda^N) + ln\left(e^{-\lambda \sum_{i=1}^{N} x_i}\right) = N ln(\lambda) - \lambda \sum_{i=1}^{N} x_i$
- $\frac{\partial ln\big(L(\lambda)\big)}{\partial \lambda} = 0 \Longrightarrow N\frac{1}{\lambda} - \sum_{i=1}^{N} x_i = 0 \Longrightarrow \lambda = \frac{N}{\sum_{i=1}^{N} x_i} \Longrightarrow \lambda = \frac{1}{\bar{x}} = \frac{1}{E[x]}$

Exemplo com distribuição Exponencial

Amostra com 5 elementos {4; 3; 2; 1; 3} → $\bar{x}$ = 2,6



Exercício

Considere a fdp abaixo. Construa um estimador de máxima verossimilhança para o parâmetro λ.

$$f(x) = \begin{cases} \dfrac{x^2 e^{\left(-x/\lambda\right)}}{2\lambda^3};\ \ x > 0 \\ 0; caso\ contrário \end{cases}$$

Exercício

Considere a fdp abaixo. Construa um estimador de máxima verossimilhança para o parâmetro λ.

$$f(x) = \begin{cases} \frac{\theta+3}{3} x^{\frac{\theta}{3}}; & 0 \leq x \leq 1 \ e \ \theta > 0 \\ 0; & caso \ contrário \end{cases}$$

Estimativa de parâmetros

- Estimativa por ponto
  - Uma amostras
  - Várias amostras – média ponderada
- Estimativa por intervalo de confiança (IC)

  O IC é construído de tal forma que se o processo de construção for repetido um grande número de vezes, o percentual de IC que contém de fato o parâmetro estimado é dado por (1- α)

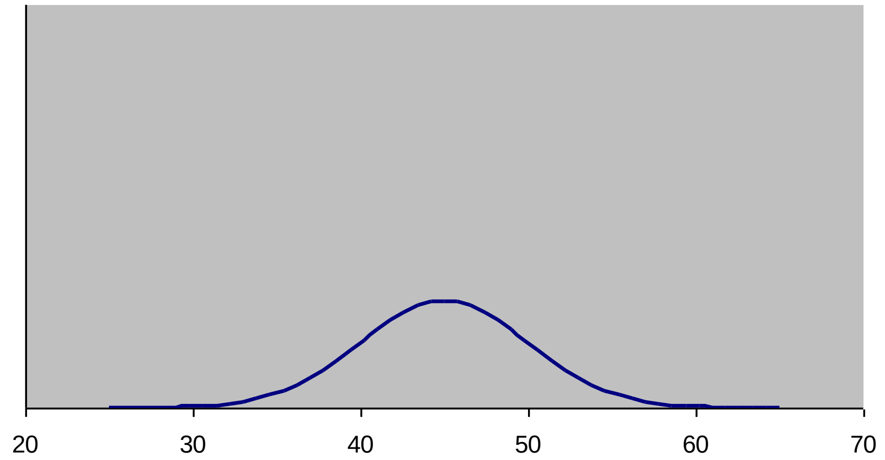
  Confiança = (1- α)

Intervalo de Confiança (IC) para:

- Média populacional – μ
- Variância populacional – σ^2
- Fração da população – p

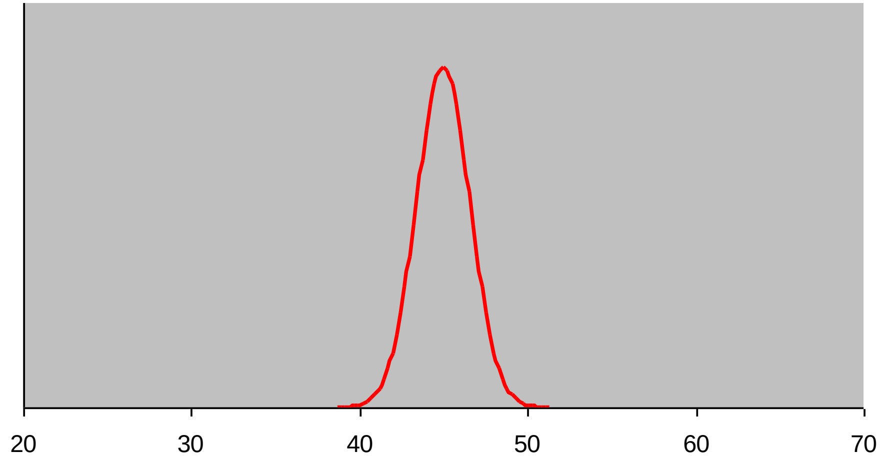
Intervalo de Confiança (IC) para média populacional – μ

Variável **x** na população

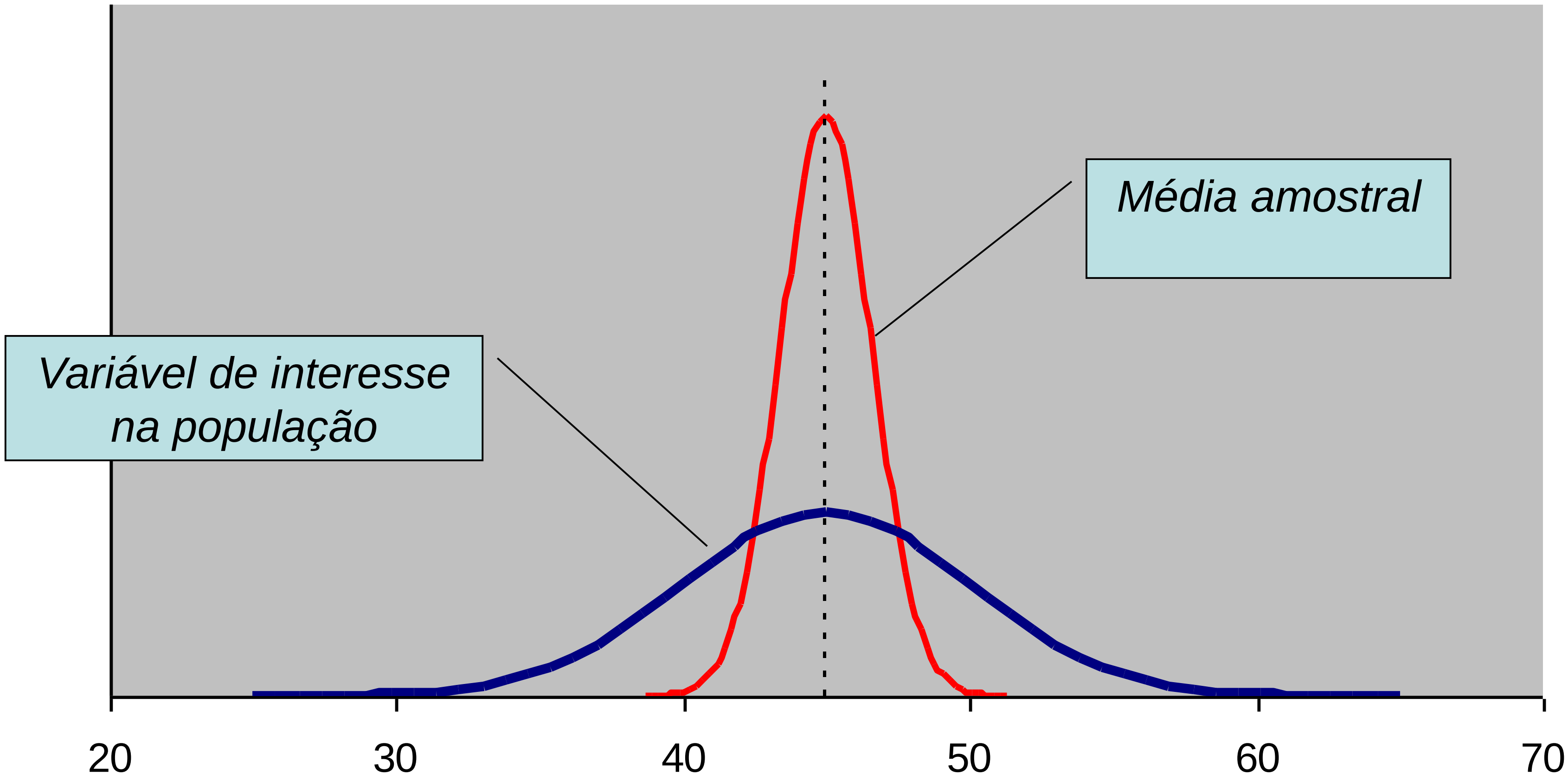


Variável média amostral de uma amostra com N elementos

$$\bar{x} = N\left(\mu; \frac{\sigma^2}{N}\right)$$



Intervalo de Confiança (IC) para média populacional – μ



Intervalo de Confiança (IC) para média populacional – μ

O limite inferior (L_i) do IC é tal que:

$P(T > t \mid \theta = L_i) = \alpha/2$

O limite superior (L_s) do IC é tal que:

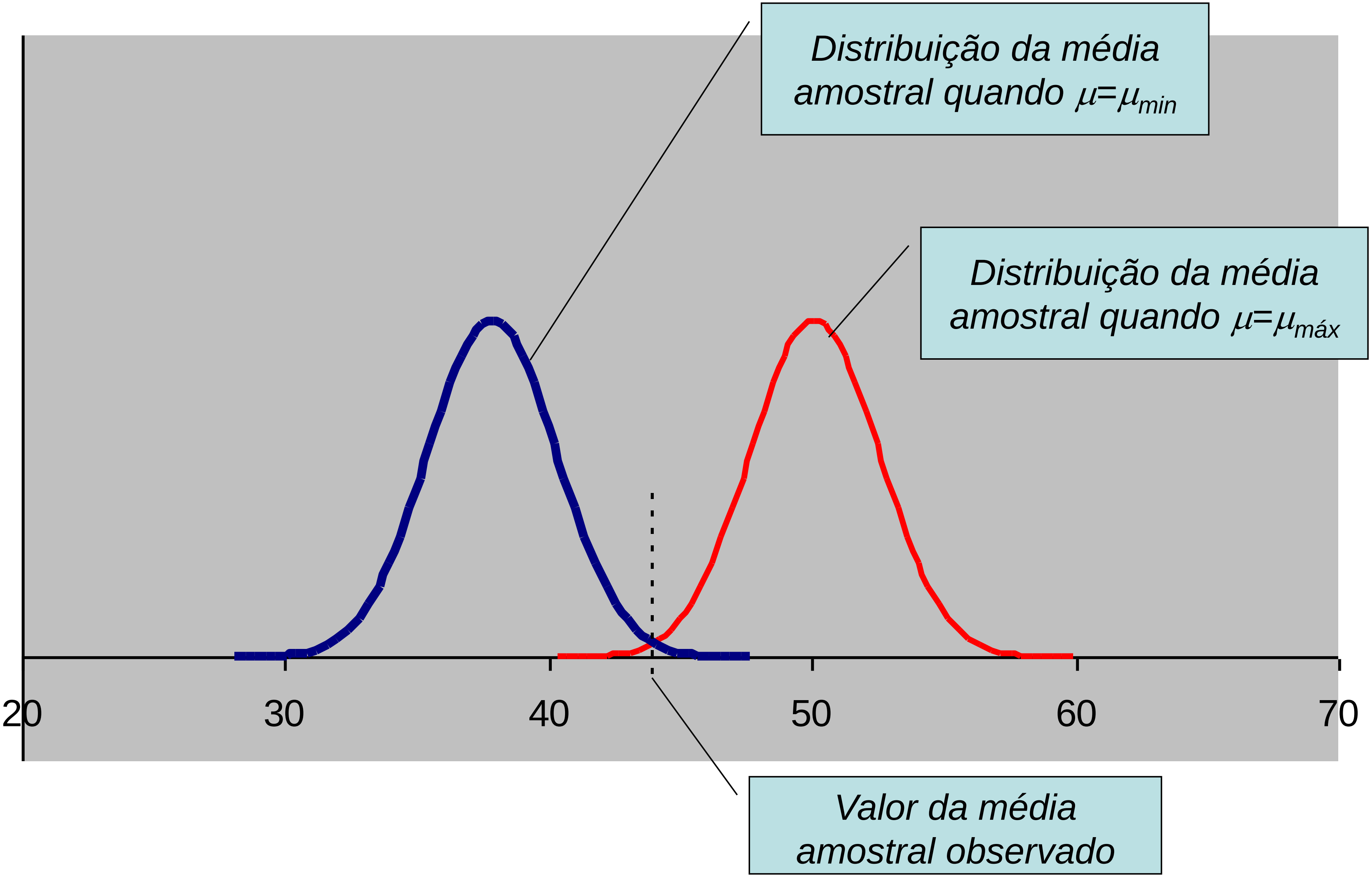
$P(T < t \mid \theta = L_s) = \alpha/2$

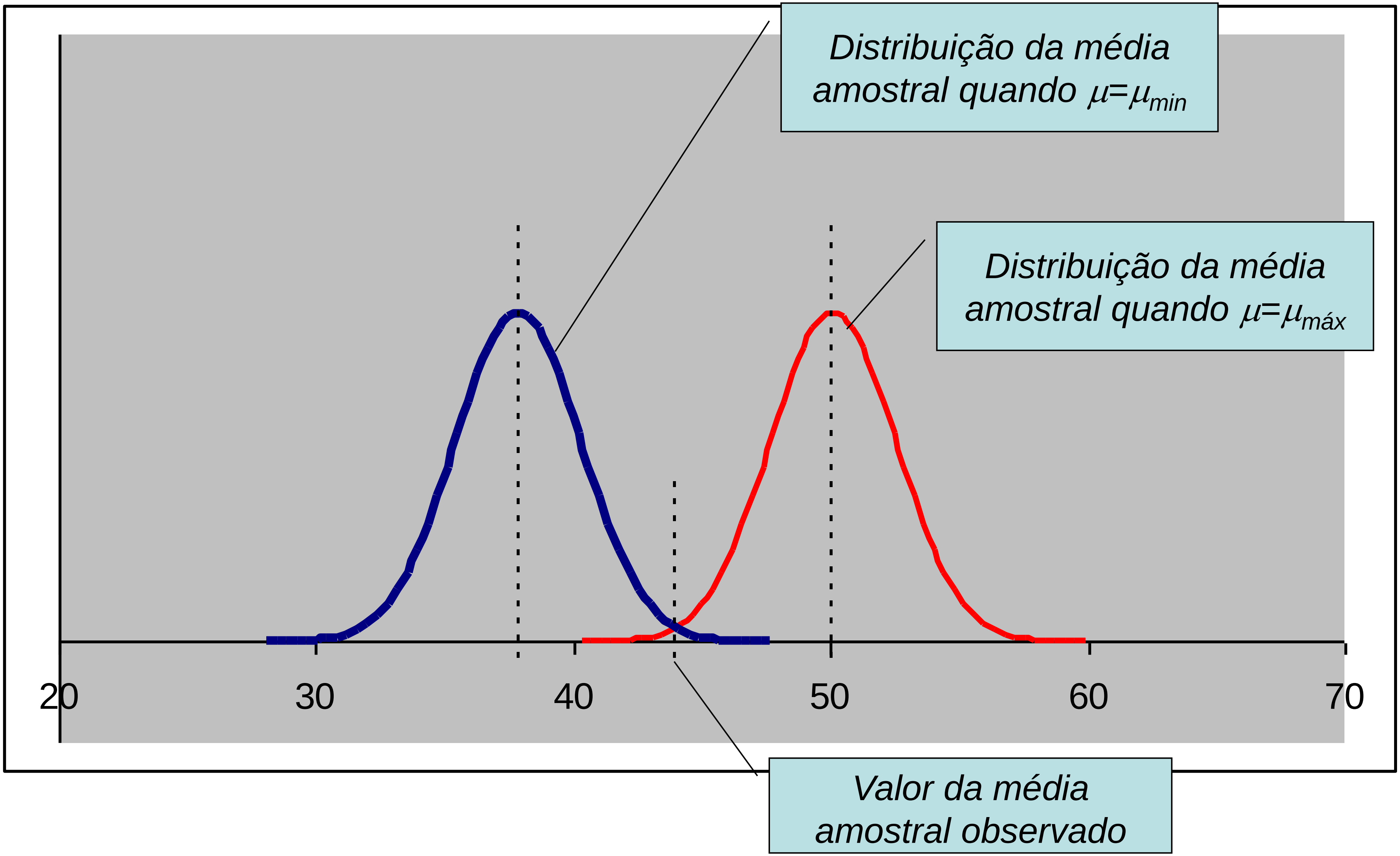
Intervalo de Confiança (IC) para média populacional – μ

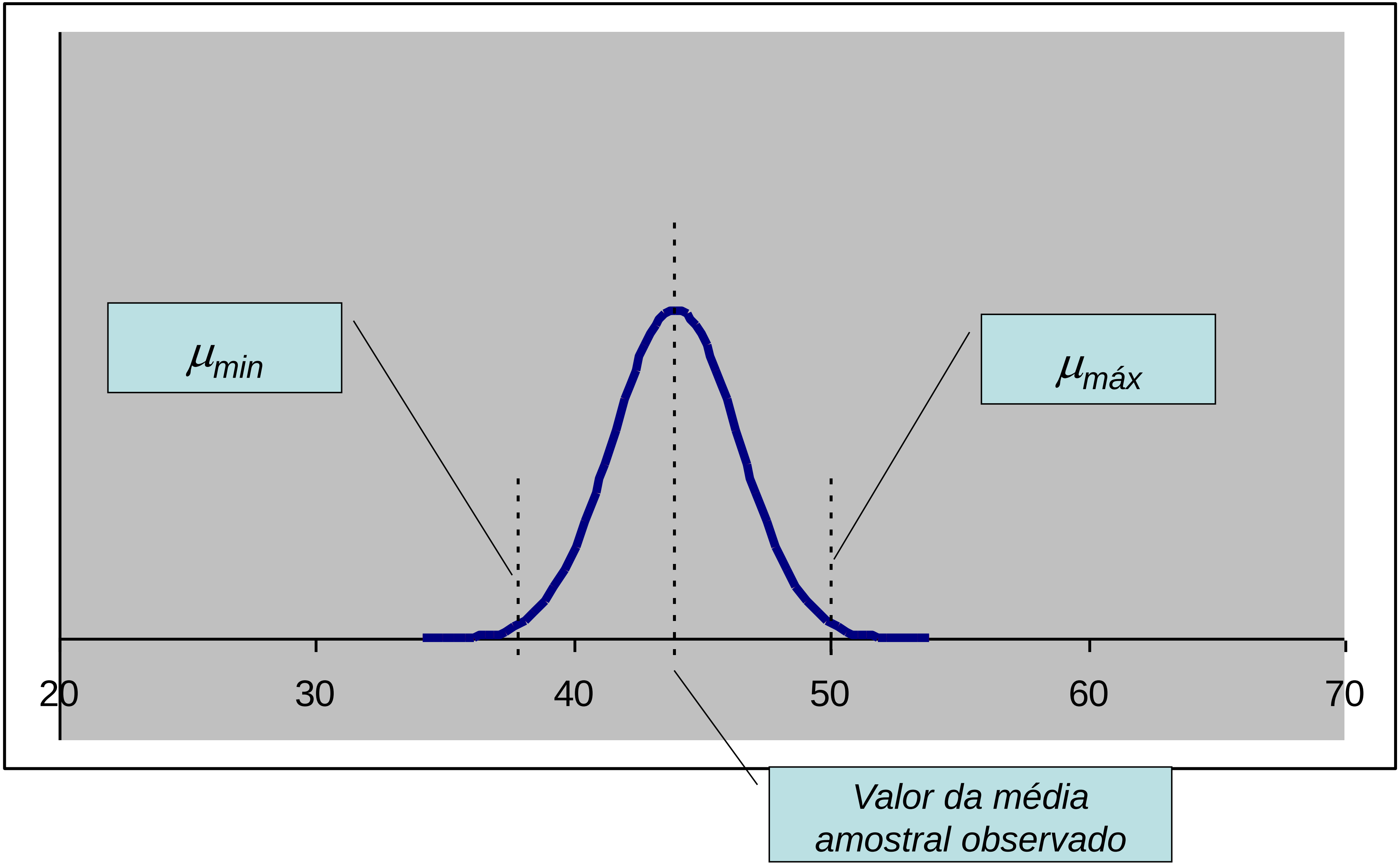
$$P\left(\bar{x} > a \mid \mu = \mu_{\min}\right) = \alpha/2$$

$$P\left(\bar{x} < a \mid \mu = \mu_{máx}\right) = \alpha/2$$

$$IC : \mu_{\min} \leq \mu \leq \mu_{máx}$$


Distribuição da média amostral quando μ=μ_{min}
Distribuição da média amostral quando μ=$\mu_{máx}$
20
30
40
50
60
70
Valor da média amostral observado

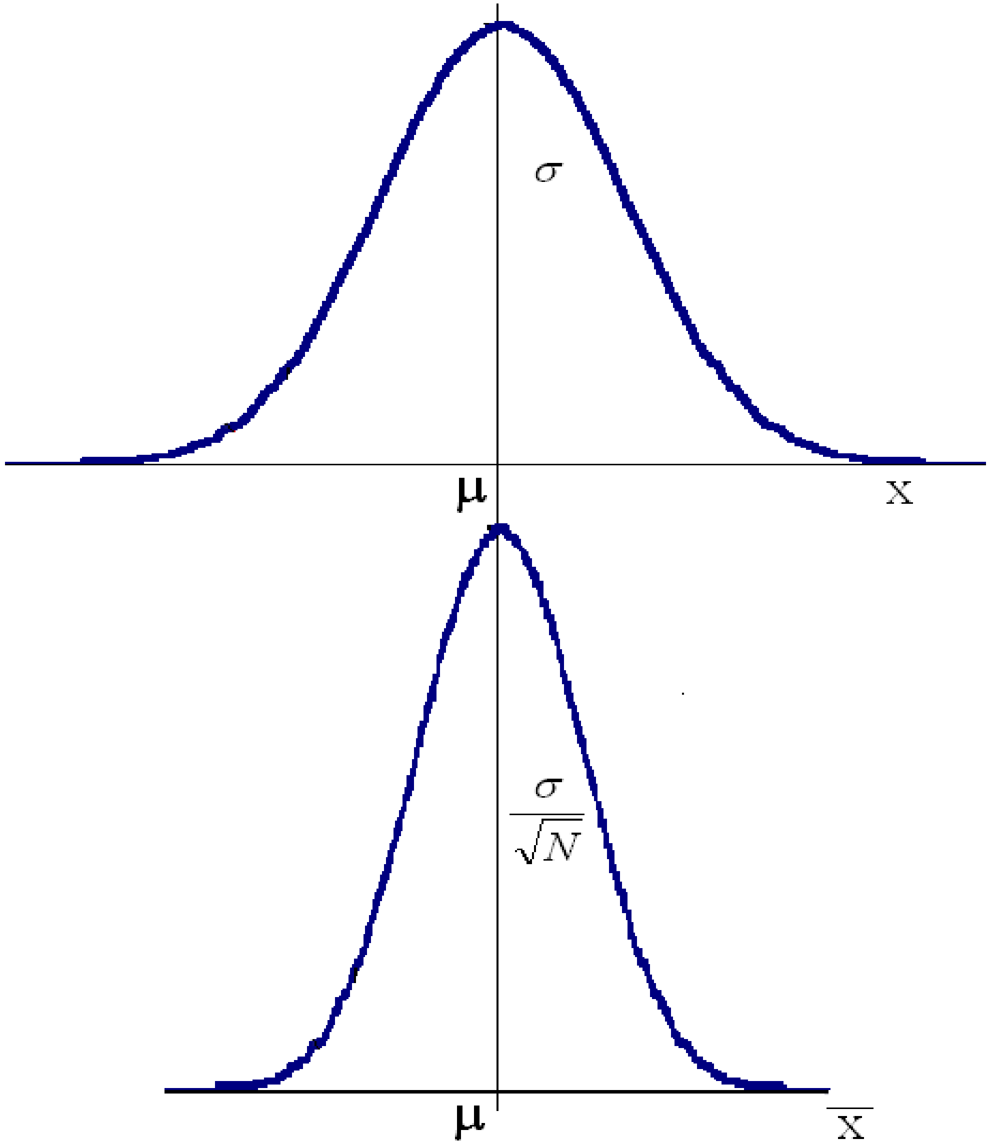

Distribuição da média amostral quando $\mu=\mu_{min}$
Distribuição da média amostral quando $\mu=\mu_{máx}$
20
30
40
50
60
70
Valor da média amostral observado


μ_{min}
$\mu_{máx}$
20
30
40
50
60
70
Valor da média amostral observado

Intervalo de Confiança para Média Populacional

Variância Conhecida

$$\mu = \bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{N}}$$



Exemplo

Suponha que uma amostra com 16 elementos teve uma média de 13,6. Construa um IC para a média populacional com nível de confiança de 90%. Suponha que o desvio padrão populacional (σ) é de 2,1.

$$\sigma = 2{,}1$$
$$N = 16$$
$$\bar{x} = 13{,}6$$
$$1 - \alpha = 90\%$$
$$z_{\alpha/2} = 1{,}65$$

$$\mu = \bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{N}}$$

$$\mu = 13{,}6 \pm 1{,}65 \times \frac{2{,}1}{\sqrt{16}}$$

$$\mu = 13{,}6 \pm 0{,}9$$

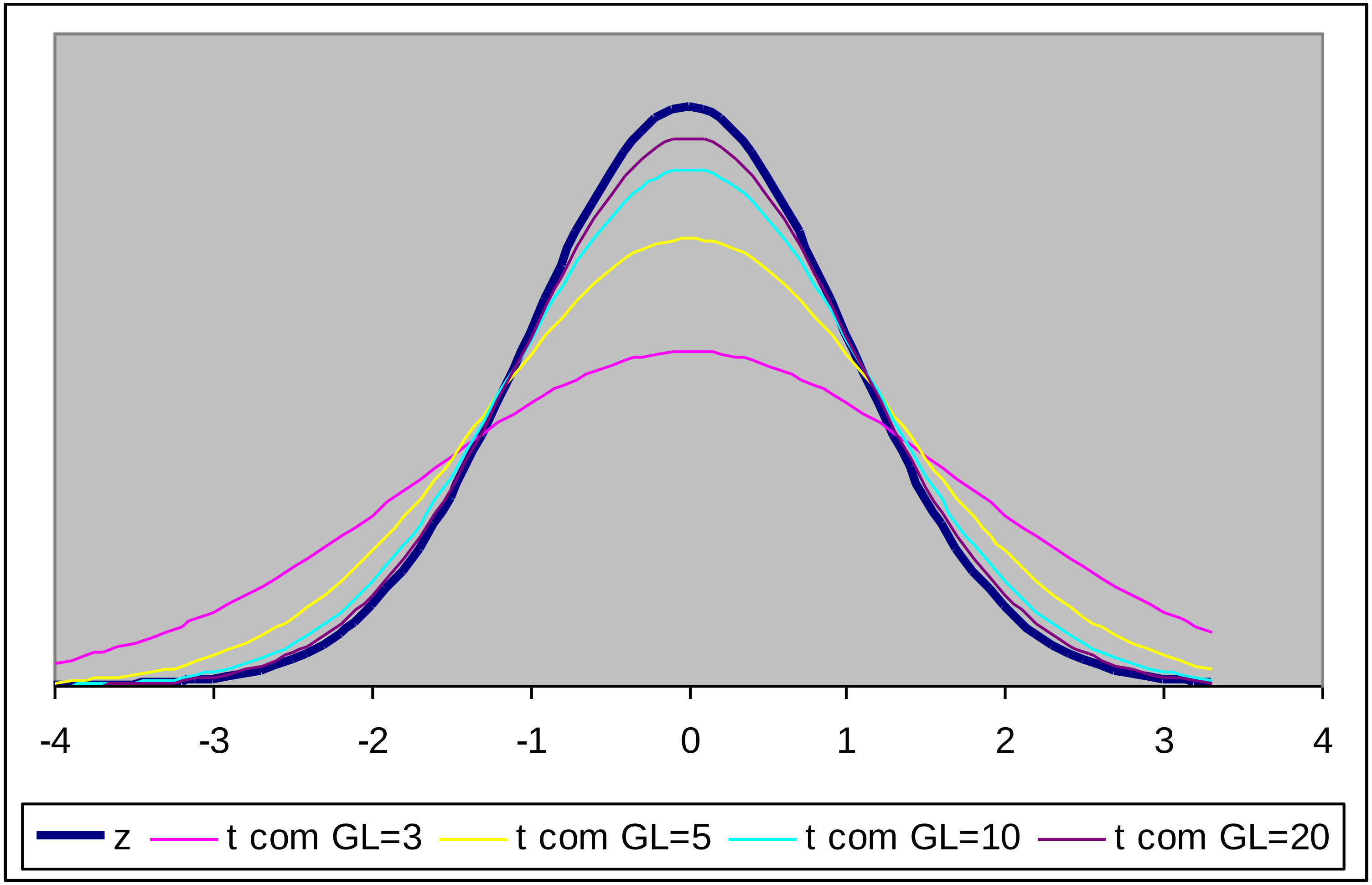
$$12{,}7 \le \mu \le 14{,}5$$

Intervalo de Confiança para Média Populacional

Variância Desconhecida

$$\mu = \bar{x} \pm t_{\alpha/2;N-1} \frac{s}{\sqrt{N}}$$

Distribuição t de Student



Distribuição de t de Student

$$\mu = 0$$

$$\sigma^2 = \frac{\nu}{\nu - 2}$$

$$\nu : \text{Graus de Liberdade}$$

$$\nu = N - 1$$

Student é um pseudônimo de William Sealy Gosset, que não podia publicar artigos usando seu próprio nome.

Exemplo

Suponha que uma amostra com 16 elementos teve uma média de 13,6 e desvio padrão 2,1. Construa um IC para a média populacional com nível de confiança de 90%.

$$N = 16$$

$$\nu = N - 1 = 15$$

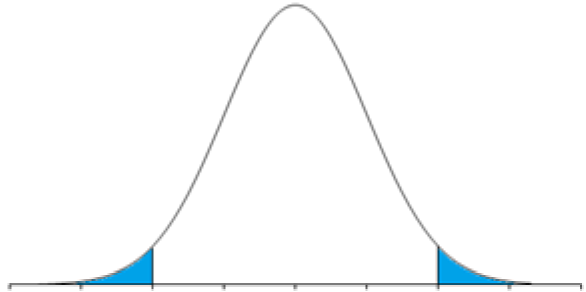
$$\bar{x} = 13{,}6$$

$$s = 2{,}1$$

$$1 - \alpha = 90\%$$

$$\alpha = 10\%$$

$$t_{\alpha/2\,N-1} = 1{,}753$$



Função t de Student

ν Graus de liberdade	α (bi-caudal)										
	50,0%	25,0%	20,0%	10,0%	5,0%	2,5%	2,0%	1,0%	0,5%	0,3%	0,1
	α (mono-caudal)										
	25,0%	12,5%	10,0%	5,0%	2,5%	1,3%	1,0%	0,5%	0,3%	0,1%	0,05
1	1,00	2,41	3,08	6,31	12,71	25,45	31,82	63,66	127,32	254,65	636,6
2	0,816	1,604	1,886	2,920	4,303	6,205	6,965	9,92	14,09	19,96	31,6
3	0,765	1,423	1,638	2,353	3,182	4,177	4,541	5,84	7,45	9,46	12,9
4	0,741	1,344	1,533	2,132	2,776	3,495	3,747	4,604	5,598	6,758	8,61
5	0,727	1,301	1,476	2,015	2,571	3,163	3,365	4,032	4,773	5,604	6,86
6	0,718	1,273	1,440	1,943	2,447	2,969	3,143	3,707	4,317	4,981	5,95
7	0,711	1,254	1,415	1,895	2,365	2,841	2,998	3,499	4,029	4,595	5,40
8	0,706	1,240	1,397	1,860	2,306	2,752	2,896	3,355	3,833	4,334	5,04
9	0,703	1,230	1,383	1,833	2,262	2,685	2,821	3,250	3,690	4,146	4,78
10	0,700	1,221	1,372	1,812	2,228	2,634	2,764	3,169	3,581	4,005	4,58
11	0,697	1,214	1,363	1,796	2,201	2,593	2,718	3,106	3,497	3,895	4,43
12	0,695	1,209	1,356	1,782	2,179	2,560	2,681	3,055	3,428	3,807	4,31
13	0,694	1,204	1,350	1,771	2,160	2,533	2,650	3,012	3,372	3,735	4,22
14	0,692	1,200	1,345	1,761	2,145	2,510	2,624	2,977	3,326	3,675	4,14
15	0,691	1,197	1,341	1,753	2,131	2,490	2,602	2,947	3,286	3,624	4,07
16	0,690	1,194	1,337	1,746	2,120	2,473	2,583	2,921	3,252	3,581	4,01
17	0,689	1,191	1,333	1,740	2,110	2,458	2,567	2,898	3,222	3,543	3,96
18	0,688	1,189	1,330	1,734	2,101	2,445	2,552	2,878	3,197	3,510	3,92
19	0,688	1,187	1,328	1,729	2,093	2,433	2,539	2,861	3,174	3,481	3,88

Dados da amostra

$$N = 16$$

$$\bar{x} = 13,6$$

$$s = 2,1$$

$$1 - \alpha = 90\%$$

$$\alpha = 10\%$$

$$\left.\begin{array}{l} N = 16 \rightarrow \nu = 15 \\ 1 - \alpha = 90\% \rightarrow \alpha = 10\% \end{array}\right\} \Rightarrow t_{\nu;\alpha/2} = t_{15;5\%} = 1,753$$

$$\mu = \bar{x} \pm t_{\nu;\alpha/2} \frac{s}{\sqrt{N}} = 13,6 \pm 1,753 \times \frac{2,1}{\sqrt{16}} = 13,6 \pm 0,92$$

$$\mu = 13,6 \pm 0,92$$

$$12,68 \leq \mu \leq 14,52$$

Dimensionamento do tamanho da amostra

Qual deve ser o tamanho da amostra de forma que a semi amplitude do intervalo de confiança não seja superior a 0,6 e o nível de confiança seja de 99,0%?

Dado que o desvio padrão (e a variância) populacional é conhecida.(σ = 2,1)

Dimensionamento do tamanho da amostra- σ^2 conhecido

$$\mu = \bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{N}}$$

$$e = z_{\alpha/2} \frac{\sigma}{\sqrt{N}}$$

$$z_{\alpha/2} \frac{\sigma}{\sqrt{N}} \le e \Rightarrow \longrightarrow \boxed{N \ge \left(\frac{z_{\alpha/2}\sigma}{e} \right)^2}$$

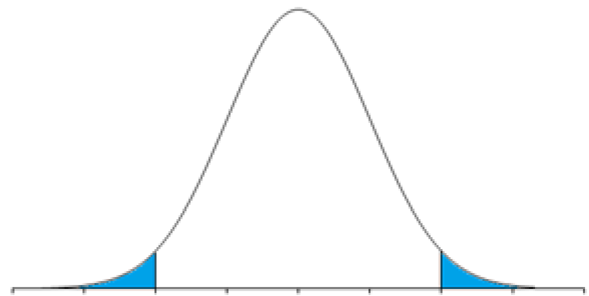
$$N \ge \left(\frac{2{,}58 \times 2{,}1}{0{,}6} \right)^2 = 81{,}54$$

$$N = 82$$

Dimensionamento do tamanho da amostra

Qual deve ser o tamanho da amostra de forma que a semi amplitude do intervalo de confiança não seja superior a 0,6 e o nível de confiança seja de 99,0%?

Neste caso o desvio padrão (e a variância) populacional NÃO é conhecido. Assim, uma amostra piloto (N=16) será utilizada para se obter uma estimativa (s=2,1) do desvio padrão da população.



ν Graus de liberdade	Função t de Student									
	α (bi-caudal)									
	50,0%	25,0%	20,0%	10,0%	5,0%	2,5%	2,0%	1,0%	0,5%	0,3%
	α (mono-caudal)									
	25,0%	12,5%	10,0%	5,0%	2,5%	1,3%	1,0%	0,5%	0,3%	0,1%
1	1,00	2,41	3,08	6,31	12,71	25,45	31,82	63,66	127,32	254,65
2	0,816	1,604	1,886	2,920	4,303	6,205	6,965	9,92	14,09	19,96
3	0,765	1,423	1,638	2,353	3,182	4,177	4,541	5,84	7,45	9,46
4	0,741	1,344	1,533	2,132	2,776	3,495	3,747	4,604	5,598	6,758
5	0,727	1,301	1,476	2,015	2,571	3,163	3,365	4,032	4,773	5,604
6	0,718	1,273	1,440	1,943	2,447	2,969	3,143	3,707	4,317	4,981
7	0,711	1,254	1,415	1,895	2,365	2,841	2,998	3,499	4,029	4,595
8	0,706	1,240	1,397	1,860	2,306	2,752	2,896	3,355	3,833	4,334
9	0,703	1,230	1,383	1,833	2,262	2,685	2,821	3,250	3,690	4,146
10	0,700	1,221	1,372	1,812	2,228	2,634	2,764	3,169	3,581	4,005
11	0,697	1,214	1,363	1,796	2,201	2,593	2,718	3,106	3,497	3,895
12	0,695	1,209	1,356	1,782	2,179	2,560	2,681	3,055	3,428	3,807
13	0,694	1,204	1,350	1,771	2,160	2,533	2,650	3,012	3,372	3,735
14	0,692	1,200	1,345	1,761	2,145	2,510	2,624	2,977	3,326	3,675
15	0,691	1,197	1,341	1,753	2,131	2,490	2,602	2,947	3,286	3,624
16	0,690	1,194	1,337	1,746	2,120	2,473	2,583	2,921	3,252	3,581
17	0,689	1,191	1,333	1,740	2,110	2,458	2,567	2,898	3,222	3,543

Dimensionamento do tamanho da amostra - σ^2 desconhecido

$$\mu = \bar{x} \pm t_{N-1;\alpha/2} \frac{s}{\sqrt{N}}$$

$$e = t_{N-1;\alpha/2} \frac{s}{\sqrt{N}}$$

$$t_{N-1;\alpha/2} \frac{s}{\sqrt{N}} \leq e \Rightarrow \longrightarrow \boxed{N \geq \left(\frac{t_{N-1;\alpha/2}\, s}{e} \right)^2}$$

$$N \geq \left(\frac{2{,}947 \times 2{,}1}{0{,}6} \right)^2 = 106{,}29$$

$$N = 107$$

Intervalo de Confiança para Fração da população – p

- Distribuição Binomial
- Aproximação pela Normal

Aproximação pela Normal

Condição Necessária

Np>5

N(1-p)>5

Distribuição Binomial

N: Número de provas de Bernoulli

P: probabilidade de sucesso mas provas de Bernoulli

X: número de sucessos

E[X]=Np

Var[X]=NPq=Np(1-p)

Distribuição Normal

E[X]=Np

Var[X]=Npq=Np(1-p)

Proporção amostral

Feita a aproximação pela normal a proporção amostral será dada por

$$\hat{p} = \frac{X}{N}$$

$$E[\hat{p}] = E\left[\frac{X}{N}\right] = \frac{1}{N}E[X] = \frac{1}{N}Np = p$$

$$Var[\hat{p}] = \frac{1}{N^2}Var[X] = \frac{1}{N^2}Npq = \frac{pq}{N}$$

$$\text{Desvio Padrão}[\hat{p}] = \sqrt{\frac{pq}{N}}$$

Exercício 45

Foi feita uma pesquisa eleitoral no bairro A com 500 eleitores sendo que 100 deles manifestaram a intenção de votar no candidato H. No bairro B, foram entrevistados 1000 eleitores e 300 deles manifestaram interesse em votar em H.

a) Construa o intervalo de confiança para a probabilidade de votar em H no bairro A (confiança: 95%).

b) Construa o intervalo de confiança para a probabilidade de votar em H no bairro B (confiança: 95%).

c) Construa o intervalo de confiança para a diferença entre as probabilidades de votar em H no bairro A e B (confiança: 95%).

Exercício 45

Bairro A

$$\mu = p$$

$$\sigma = \sqrt{\frac{pq}{N}}$$

$$p = \frac{100}{500} = 20\%$$

$$\mu = 20\%$$

$$\sigma = 1,8\%$$

$$IC:$$

$$p = \widehat{p} \pm z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{N}}$$

$$p = 20\% \pm 3,5\%$$

Bairro B

$$\mu = p$$

$$\sigma = \sqrt{\frac{pq}{N}}$$

$$p = \frac{300}{1000} = 30\%$$

$$\mu = 30\%$$

$$\sigma = 1,4\%$$

$$IC:$$

$$p = \widehat{p} \pm z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{N}}$$

$$p = 30\% \pm 2,8\%$$

Exercício 45

$$\mu_A = 20\%$$
$$\sigma_A = 1{,}8\%$$
$$\mu_B = 30\%$$
$$\sigma_B = 1{,}4\%$$
$$X = p_B - p_A$$
$$\mu_X = \mu_B - \mu_A = 10\%$$
$$\sigma_X = \sqrt{(\sigma_B)^2 + (\sigma_A)^2} = 2{,}3\%$$

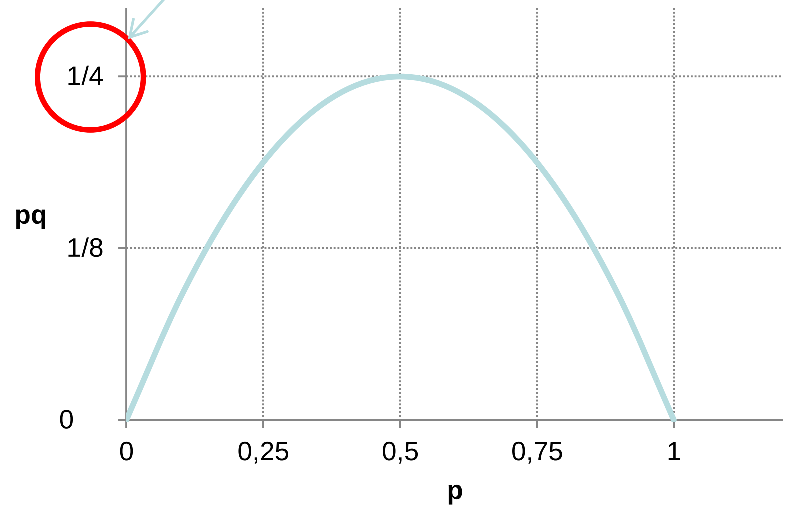
$$IC:$$
$$\mu_X = 10\% \pm z_{\alpha/2}\, 2{,}3\%$$
$$\mu_X = 10\% \pm 1{,}96 \times 2{,}3\%$$
$$\mu_X = 10\% \pm 4{,}51\%$$

Dimensionamento do tamanho da amostra

Estimativa de uma fração da população

Qual deve ser o tamanho da amostra de forma que a semi amplitude do intervalo de confiança não seja superior a 1% e o nível de confiança seja de 90%?

Máximo valor de ***pq***



$$\hat{p} = p \pm z_{\alpha/2}\sqrt{\frac{pq}{N}}$$

$$e = z_{\alpha/2}\sqrt{\frac{pq}{N}}$$

$$z_{\alpha/2}\sqrt{\frac{pq}{N}} \le e \longrightarrow N \ge pq\left(\frac{z_{\alpha/2}}{e}\right)^2 \longrightarrow N \ge \left(\frac{z_{\alpha/2}}{2e}\right)^2$$

$$N \ge \left(\frac{1{,}64}{2\times 0{,}01}\right)^2 = 82^2 = 6.724$$

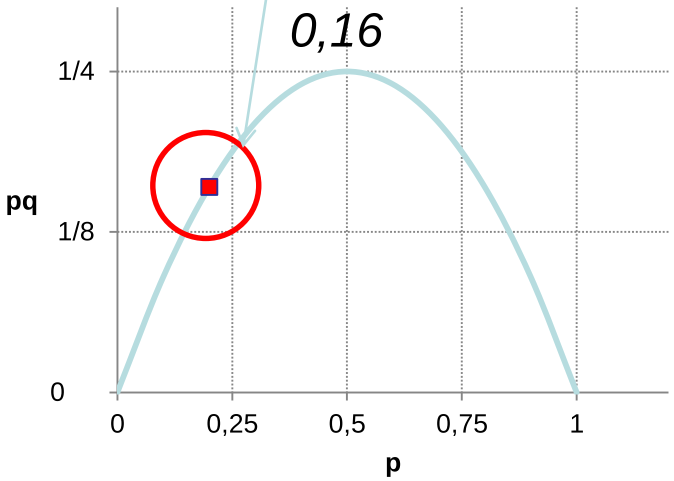
Dimensionamento do tamanho da amostra

Estimativa de uma fração da população

Qual deve ser o tamanho da amostra de forma que a semi amplitude do intervalo de confiança não seja superior a 1% e o nível de confiança seja de 90%?

É sabido que esta fração é não superior a 20%!

*Máximo valor de **pq** é 0,16*



$$\hat{p} = p \pm z_{\alpha/2}\sqrt{\frac{pq}{N}}$$

$$e = z_{\alpha/2}\sqrt{\frac{pq}{N}}$$

$$z_{\alpha/2}\sqrt{\frac{pq}{N}} \leq e \longrightarrow N \geq pq\left(\frac{z_{\alpha/2}}{e}\right)^2 \longrightarrow N \geq 0{,}16 \times \left(\frac{z_{\alpha/2}}{e}\right)^2$$

$$N \geq 0{,}16 \times \left(\frac{1{,}64}{0{,}01}\right)^2 = 0{,}16 \times 164^2 = 4.303{,}36$$

$$N = 4.304$$

Intervalo de Confiança para Variância populacional – σ^2

- Distribuição Qui-quadrado

Variância

Intervalo de Confiança

Distribuição Qui-quadrado (χ^2)

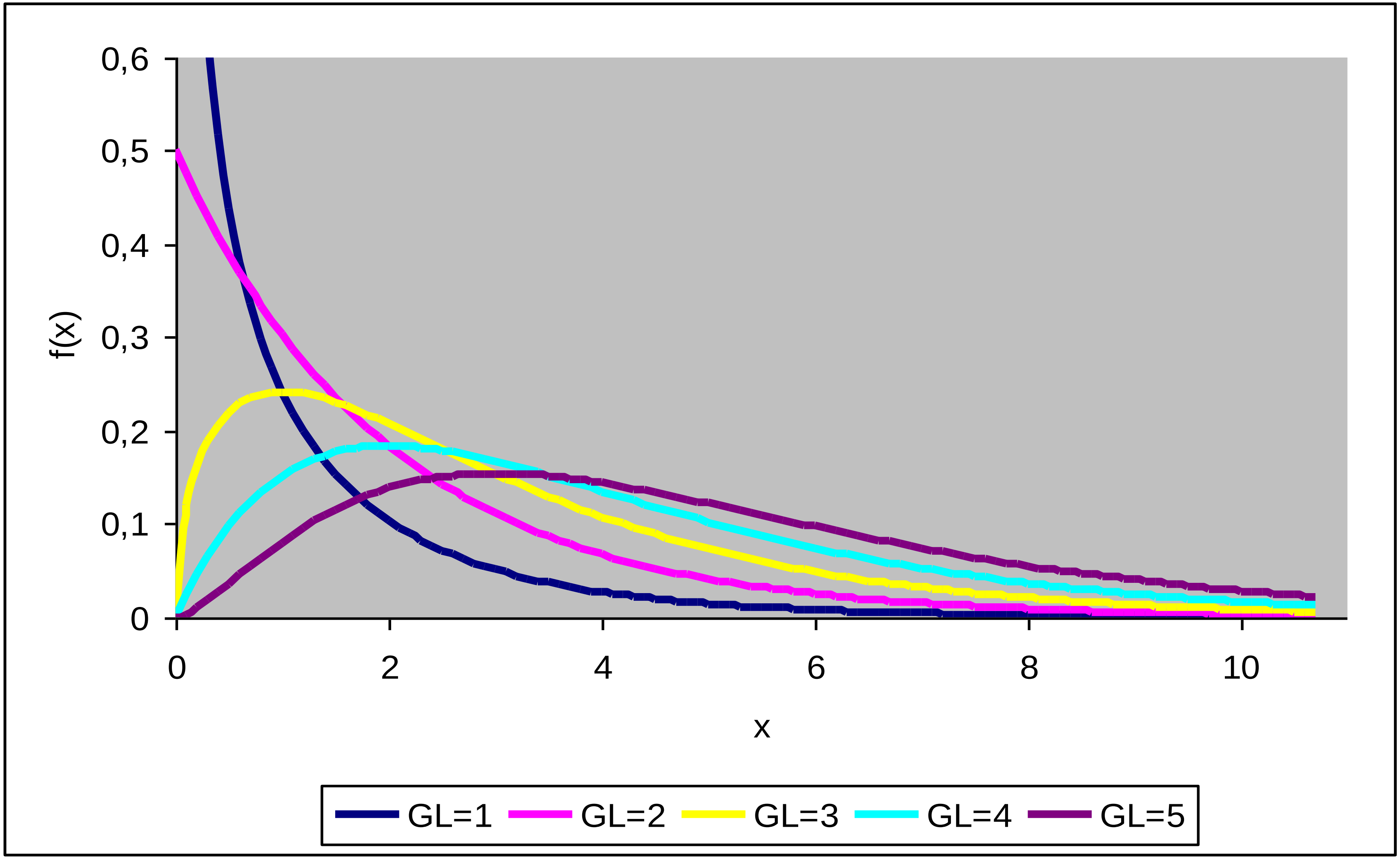
$$\chi_n^2 = \sum_{i=1}^{n} (z_i)^2$$

$$\chi_{n-1}^2 = \frac{\sum_{i=1}^{n-1} (x_i - \bar{x})^2}{\sigma^2}$$

$$\chi_{n-1}^2 = \frac{\sum_{i=1}^{n-1} (x_i - \bar{x})^2}{\sigma^2} = \frac{n-1}{\sigma^2} \times \frac{\sum_{i=1}^{n-1} (x_i - \bar{x})^2}{n-1} = \frac{n-1}{\sigma^2} \times s^2$$

$$s^2 = \frac{\sigma^2}{n-1} \times \chi_{n-1}^2$$

Distribuição Qui-quadrado χ^2



Intervalo de Confiança

$$P\left(\chi^2_{n-1;1-\alpha/2} \le \chi^2_{n-1} \le \chi^2_{n-1;\alpha/2}\right) = 1-\alpha$$

$$\chi^2_{n-1;1-\alpha/2} \le \chi^2_{n-1} \le \chi^2_{n-1;\alpha/2} = \chi^2_{n-1;1-\alpha/2} \le \frac{(n-1)s^2}{\sigma^2} \le \chi^2_{n-1;\alpha/2}$$

$$\frac{(n-1)s^2}{\chi^2_{n-1;\alpha/2}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{n-1;1-\alpha/2}}$$

Exemplo

Uma amostra de onze elementos, extraída de uma população com distribuição normal, forneceu variância $s^2 = 7,08$. Construir um intervalo de 90% de confiança para a variância desta população.

Exemplo

$$n = 11$$

$$\nu = n - 1 = 10$$

$$s^2 = 7{,}08$$

$$1 - \alpha = 90\%$$

$$\chi^2_{n-1;1-\alpha/2} = \chi^2_{10;95\%} = 3{,}9403$$

$$\chi^2_{n-1;\alpha/2} = \chi^2_{10;5\%} = 18{,}3070$$

$$\frac{(n-1)s^2}{\chi^2_{n-1;\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{n-1;1-\alpha/2}} \rightarrow \frac{10 \times 7{,}08}{18{,}3070} \leq \sigma^2 \leq \frac{10 \times 7{,}08}{3{,}9403}$$

$$3{,}87 \leq \sigma^2 \leq 17{,}97$$