ME951 - Estatística e Probabilidade I

Parte 16

1S2024

Intervalo de confiança para a média populacional μ

Revisão

- Vimos que podemos usar uma estatística para estimar um parâmetro de uma distribuição que temos interesse.
- Por ex: podemos usar a proporção amostral ($\hat{p}$), obtida a partir de um amostra aleatória, se estamos interessados na proporção populacional (p).
- Por ex: podemos usar a média amostral ($\bar{x}$), obtida a partir de um amostra aleatória, se estamos interessados na média populacional (μ).

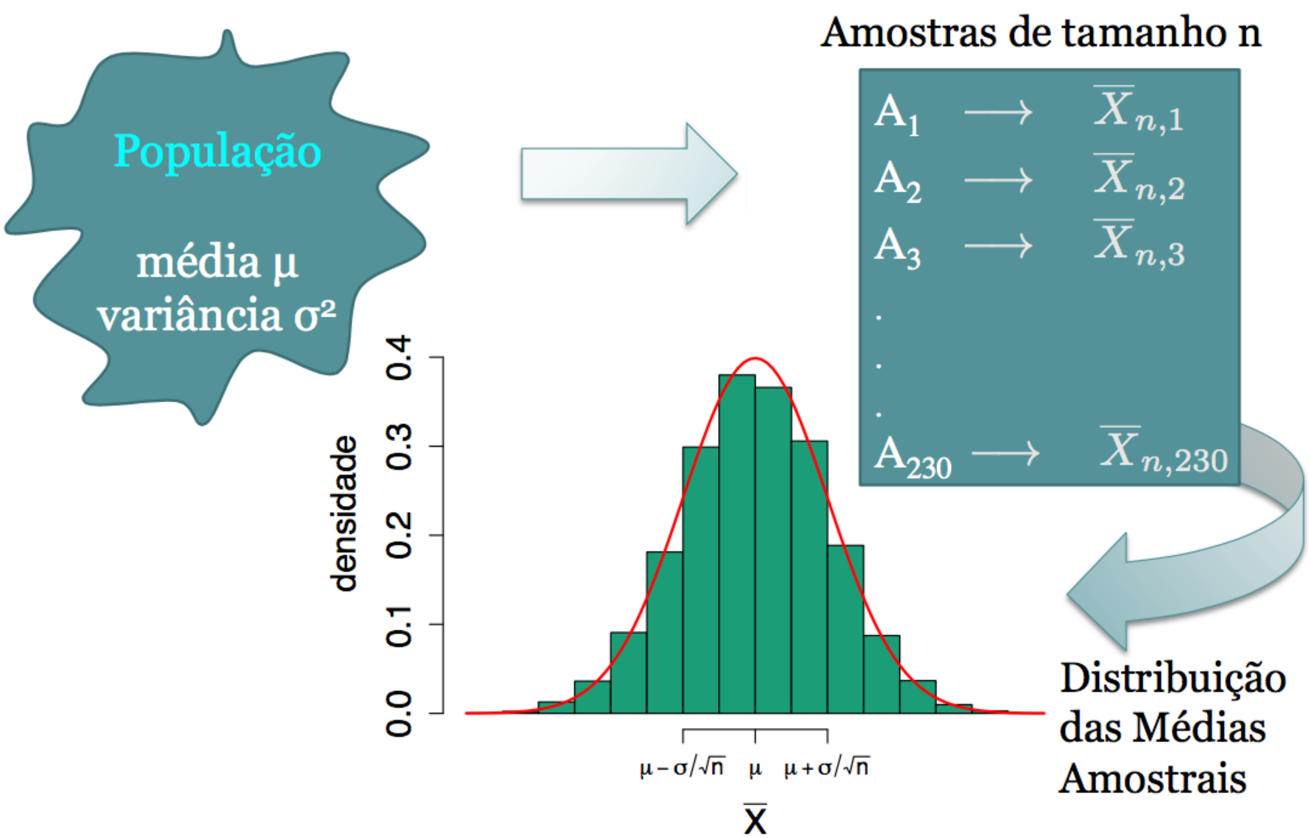
Distribuição da Média Amostral

A média amostral, $\bar{X}_n$, tem em geral valores diferentes para diferentes amostras aleatórias obtidas: é uma variável aleatória.

Para obtermos a distribuição da média amostral (empiricamente):

- Coletar uma a.a. de tamanho n a partir da população e calcular o valor da média desta amostra.
- Coletar outra a.a. de tamanho n a partir da população e calcular o valor da média desta amostra.
- Repetir isso várias vezes.
- Construir um histograma com todas as médias obtidas para estudar o comportamento de $\bar{X}_n$: avaliando a média, a dispersão e a distribuição.

Teorema Central do Limite



Teorema Central do Limite

Na prática: iremos coletar somente **uma amostra** de tamanho n. Não faremos inúmeras vezes esse processo. Com isso, teremos apenas um valor: $\bar{x}$.

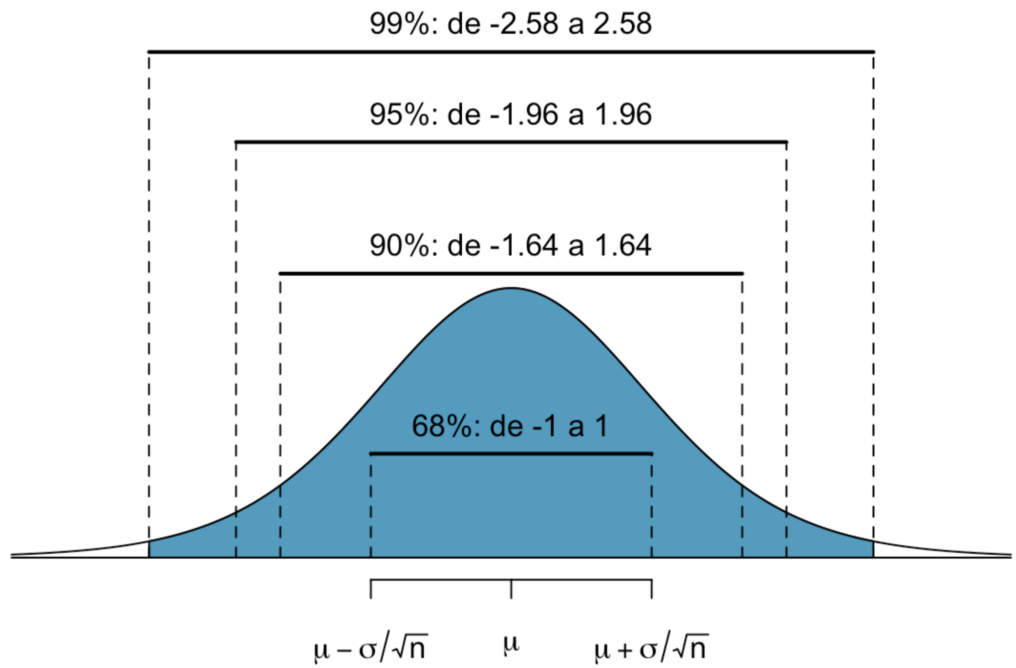
Então como saberemos as propriedades deste estimador? Quão útil ele é?

Resultado (TCL):

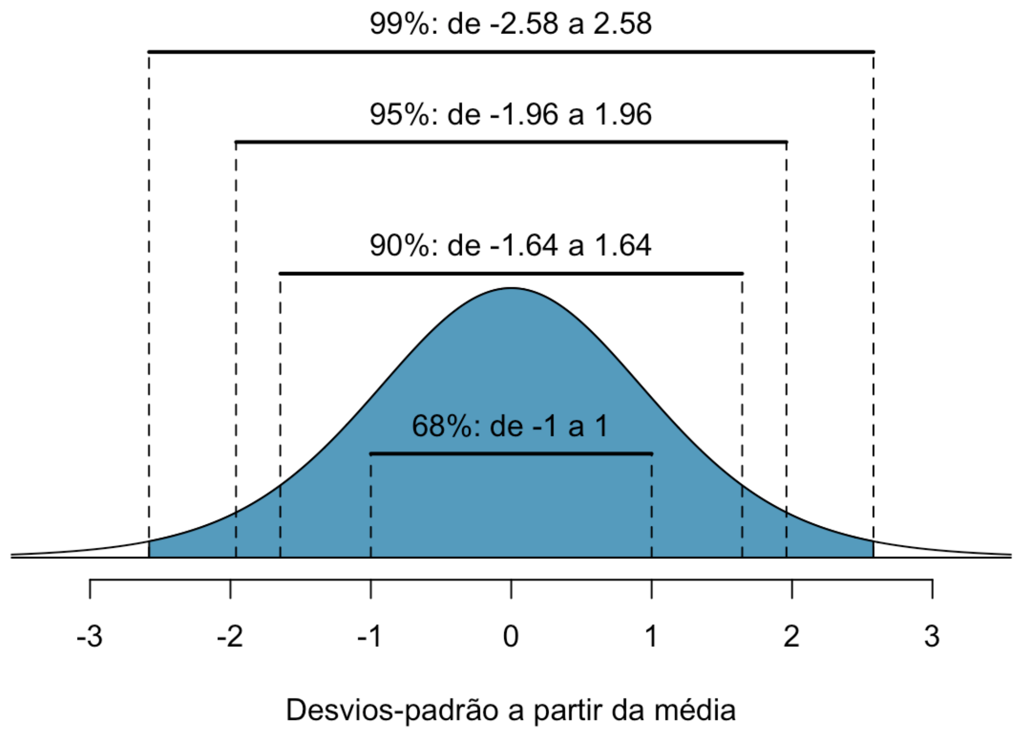
Para amostras aleatórias simples $X_1, \ldots, X_n$ coletadas de uma população com média μ e variância σ^2, a distribuição amostral de $\bar{X}_n$ aproxima-se de uma distribuição Normal:

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

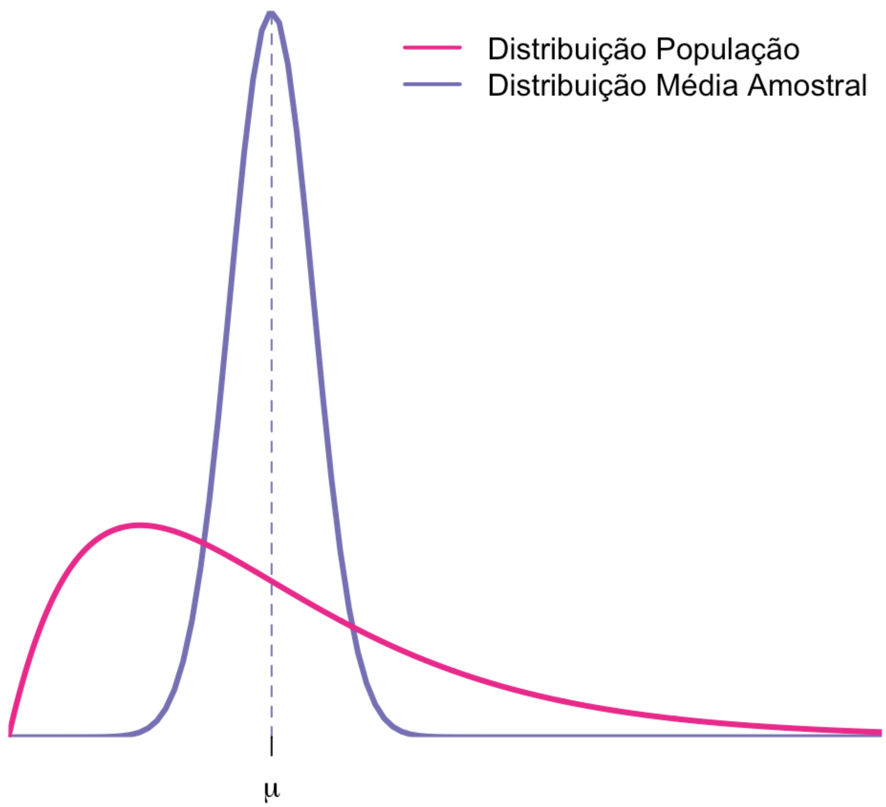
Distribuição Amostral de $\bar{X}_n$



$$Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$



Teorema Central do Limite



Qual a probabilidade de que o estimador $\bar{X}_n$ esteja distante do valor verdadeiro, μ, em no máximo 1 erro-padrão?

$$P\left(|\bar{X}_n - \mu \,|\leq \sigma/\sqrt{n}\right)$$

$$\begin{aligned}
P\left(|\bar{X}_n - \mu \,|\leq \sigma/\sqrt{n}\right) &= P\left(-\sigma/\sqrt{n} \leq \bar{X}_n - \mu \leq \sigma/\sqrt{n}\right) \\
&= P\left(-1 \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq 1\right) \\
&= P(-1 \leq Z \leq 1) \\
&= 0.68
\end{aligned}$$

Qual a probabilidade de que o estimador $\bar{X}_n$ esteja distante do valor verdadeiro, μ, em no máximo 1.96 erro-padrão?

$$P\left(|\bar{X}_n - \mu\ | \leq 1.96\,\sigma/\sqrt{n}\right)$$

$$\begin{aligned}
P\left(|\bar{X}_n - \mu\ | \leq 1.96\sigma/\sqrt{n}\right) &= P\left(-1.96\,\sigma/\sqrt{n} \leq \bar{X}_n - \mu \leq 1.96\,\sigma/\sqrt{n}\right) \\
&= P\left(-1.96 \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq 1.96\right) \\
&= P(-1.96 \leq Z \leq 1.96) \\
&= 0.95
\end{aligned}$$

Intervalo de Confiança para μ: σ conhecido

Seja $X_1, \ldots, X_n$ uma a.a. de uma população com média μ e variância σ^2 conhecida. Então,

$$Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$$

Um Intervalo de $100(1-\alpha)\%$ de confiança para μ é dado por:

$$IC(\mu, 1-\alpha) = \left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \; ; \; \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Intervalo de Confiança para μ: σ conhecido

Intervalo de confiança de 95%

$$IC(\mu, 95\%) = \left[\bar{x} - 1.96\,\sigma/\sqrt{n}; \bar{x} + 1.96\,\sigma/\sqrt{n}\right]$$

Intervalo de confiança de 90%

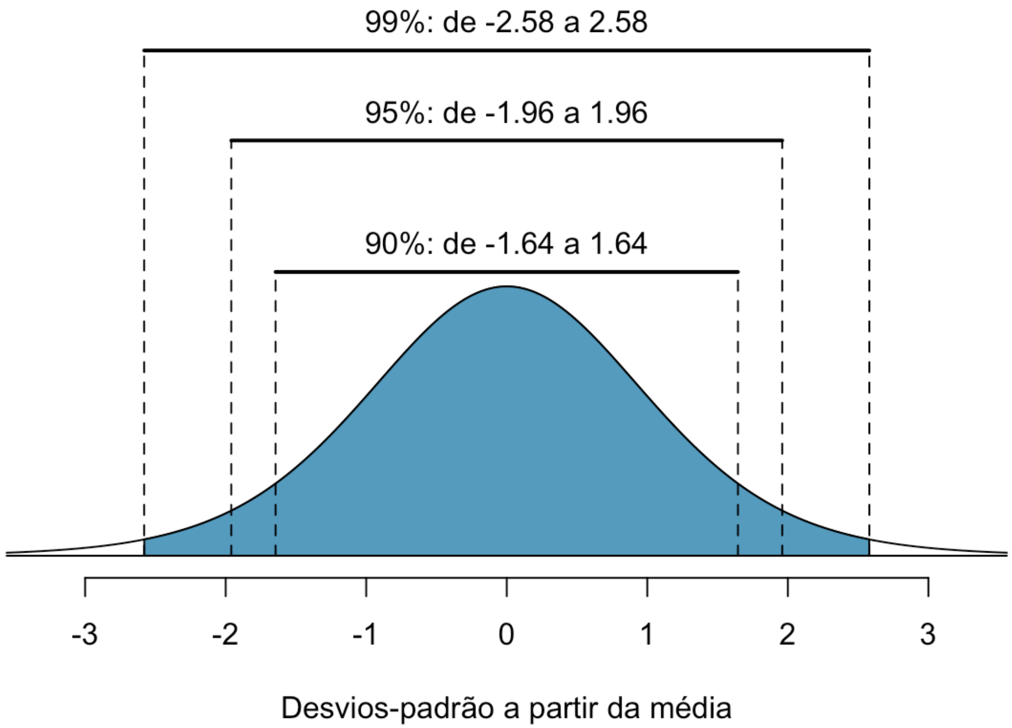
$$IC(\mu, 90\%) = \left[\bar{x} - 1.64\,\sigma/\sqrt{n}; \bar{x} + 1.64\,\sigma/\sqrt{n}\right]$$

Intervalo de confiança de 99%

$$IC(\mu, 99\%) = \left[\bar{x} - 2.58\,\sigma/\sqrt{n}; \bar{x} + 2.58\,\sigma/\sqrt{n}\right]$$

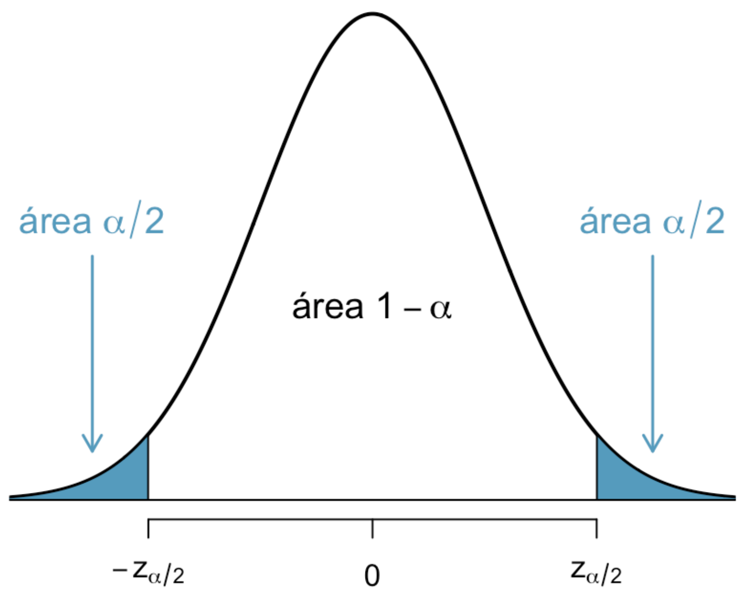
Como encontrar $z_{\alpha/2}$

$$P(|Z| \leq z_{\alpha/2}) = P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$$



Como encontrar $z_{\alpha/2}$

$$P(|Z| \leq z_{\alpha/2}) = P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$$

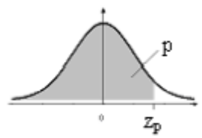


Procure na tabela o valor de z tal que a probabilidade acumulada até o valor de z, isto é $P(Z \leq z) = \Phi(z)$, seja $1 - \alpha/2$.

Exemplo

Encontrar $z_{0.05}$ tal que $0.90 = P\left(-z_{0.05} \leq Z \leq z_{0.05}\right)$.

Tabela I: Distribuição Normal Padrão Acumulada



Fornece Φ(z) = P(-∞ < Z ≤ z), para todo z, de 0,01 em 0,01, desde z = 0,00 até z = 3,59
A distribuição de Z é Normal(0;1)

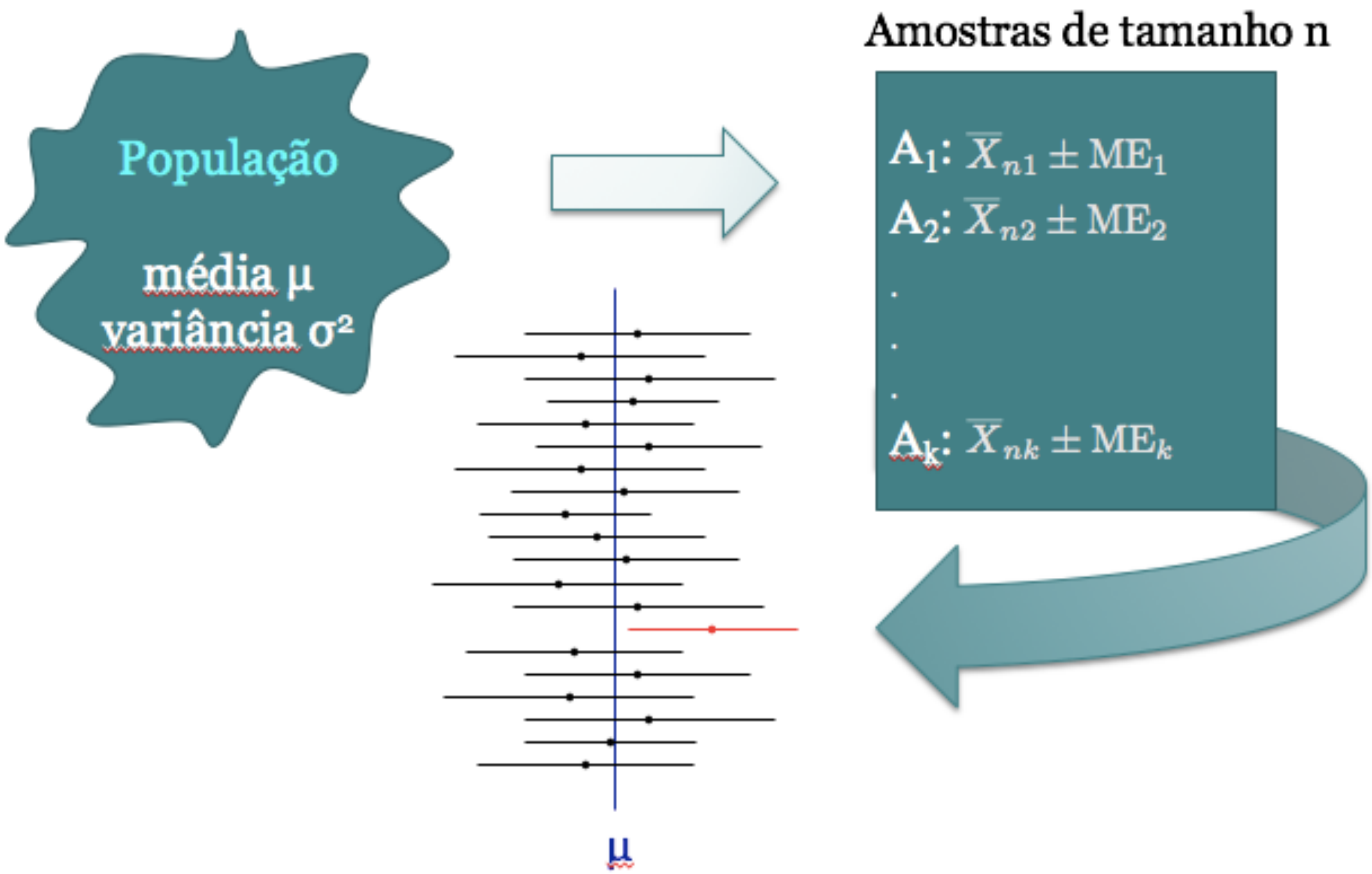
z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	$\boxed{0,9495}$	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817

Pela tabela, $z_{0.05} = 1.64$.

Interpretação do Intervalo de Confiança para μ

- Temos uma amostra aleatoria $X_1, \ldots, X_n$ e estamos usando a média amostral $\bar{X}_n$ para estimar μ, a média populacional.
- Quão boa é esta estimativa? Ela tem boa precisão? Qual o grau de confiança?
- Em geral: queremos alto grau de confiança, por exemplo, $1 - \alpha = 0.95$.
- Imagine que seja possível coletar uma amostra de tamanho n da população várias vezes. Para cada vez, você calcula $\bar{x}$ e constrói um IC de 95% para μ. Imagine também que você conhece μ e conte quantos dos intervalos contêm μ. A proporção de intervalos que contêm μ será próxima a 0.95.

Interpretação do Intervalo de Confiança para μ



Exemplo: Café

Uma máquina enche pacotes de café com variância igual a $100g^2$. Ela estava regulada para encher os pacotes com uma média de 500g. Mas o fabricante desconfia que a máquina está desregulada e quer então estimar a nova média μ.

Uma amostra de 25 pacotes apresentou uma média de 485g. Encontre um IC de 95% para a verdadeira média μ.

$\bar{x} = 485, n = 25, \sigma = 10, \alpha = 0.05, z_{0.025} = 1.96$

$$
\begin{aligned}
IC(\mu, 0.95) &= \left[\bar{x} - z_{0.025}\frac{\sigma}{\sqrt{n}};\ \bar{x} + z_{0.025}\frac{\sigma}{\sqrt{n}}\right] \\
&= \left[485 - 1.96\frac{10}{5}; 485 + 1.96\frac{10}{5}\right] \\
&= [485 - 3.92; 485 + 3.92] \\
&= [481.08; 488.92]
\end{aligned}
$$

Tamanho da Amostra

Exemplo: Por experiência, sabe-se que o peso de um salmão de certo criatório segue uma distribuição normal com uma média que varia a cada estação, mas com desvio padrão sempre igual a 0.3 libras.

Se quisermos estimar o peso médio dos peixes de maneira que nossa estimativa seja diferente da verdadeira média em no máximo 0.1 libras para mais ou para menos com probabilidade igual a 0.9, qual o tamanho amostral necessário?

$$P(|\bar{X} - \mu| \leq 0.1) = 0.9$$

Tamanho da Amostra

$$P(-0.1 \leq \bar{X} - \mu \leq 0.1) = P\left(-\frac{0.1}{\sigma/\sqrt{n}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{0.1}{\sigma/\sqrt{n}}\right)$$

$$= P\left(-\frac{0.1}{\sigma/\sqrt{n}} \leq Z \leq \frac{0.1}{\sigma/\sqrt{n}}\right) = 0.9$$

$$\frac{0.1}{\sigma/\sqrt{n}} = 1.645$$

$$n = \left(\frac{1.645\sigma}{0.1}\right)^2 = \left(\frac{1.645 \times 0.3}{0.1}\right)^2 \approx 25$$

Tamanho da Amostra

$$IC(\mu, 1-\alpha) = \left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \; ; \; \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Margem de erro: $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

Margem de erro 0.1, isto é,

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 0.1$$

$\alpha = 0.1$ (90% de confiança) e $z_{0.05} = 1.645$.

$$1.645\frac{0.3}{\sqrt{n}} = 0.1 \quad \Rightarrow \quad n = 24.35$$

Tamanho amostral: 25

Tamanho da Amostra

Em geral, para uma margem de erro m e confiança $100(1-\alpha)\%$:

$$n = \left(\frac{z_{\alpha/2}}{m}\right)^2 \sigma^2$$

Intervalo de Confiança para μ: σ desconhecido

Seja $X_1, \ldots, X_n$ uma a.a. de uma população com média μ, mas com variância σ^2 desconhecida

Nesse caso, usaremos a variância amostral (s^2) como uma estimativa de σ^2:

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

Como consequência, não temos mais distribuição Normal, mas sim a **distribuição t-student** com $n-1$ graus de liberdade:

$$T = \frac{\bar{X}_n - \mu}{\sqrt{s^2/n}} \sim t_{n-1}$$

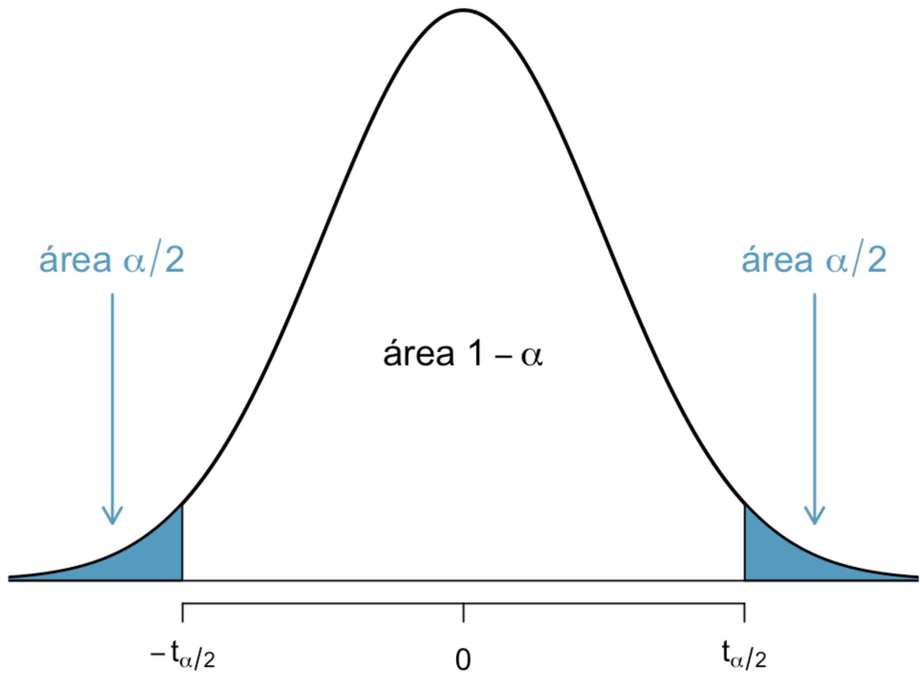
$$P(-t_{n-1,\alpha/2} < T < t_{n-1,\alpha/2}) = 1 - \alpha$$

Um intervalo de $100(1-\alpha)\%$ de confiança para μ é dado por:

$$IC(\mu, 1-\alpha) = \left[\bar{x} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}};\ \bar{x} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}\right]$$

Como encontrar $t_{n-1,\alpha/2}$

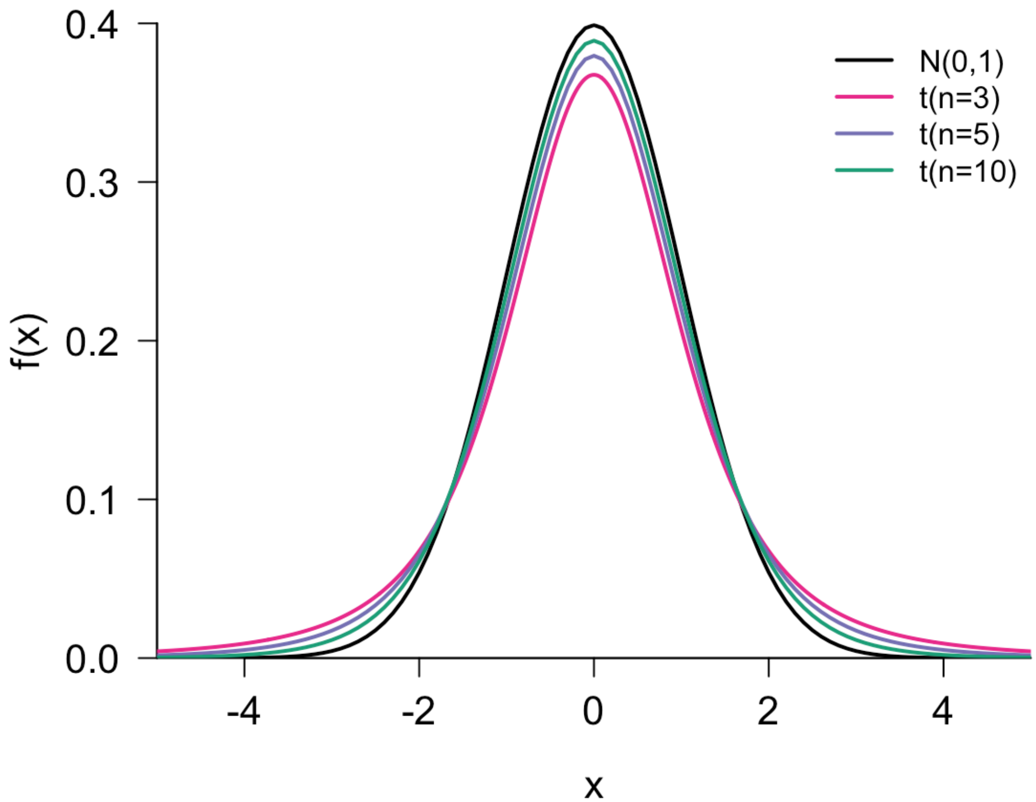
$$P(-t_{n-1,\alpha/2} < T < t_{n-1,\alpha/2}) = 1 - \alpha$$



Os valores da distribuição t-student também encontram-se tabelados.

Distribuição t-student e Normal Padrão

Para n grande a distribuição t-student se aproxima da normal padrão $N(0, 1)$.



Exemplo: Café

No exemplo da máquina que enche pacotes de café, suponha agora que a variância é **desconhecida**.

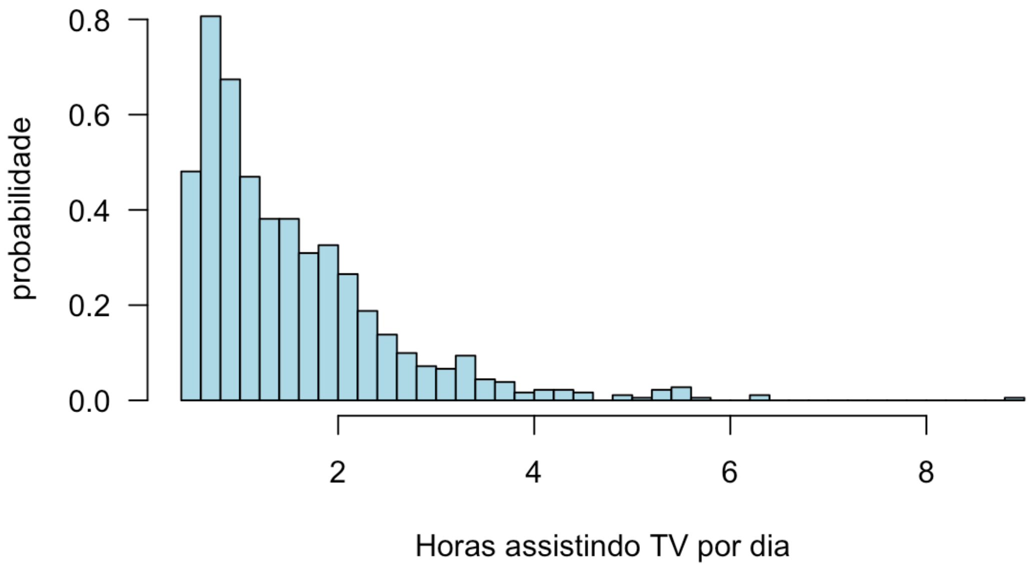
Lembre-se que uma amostra de 25 pacotes apresentou uma média de 485g. Observou-se um desvio padrão na amostra de 7.1g

Encontre um IC de 95% para a verdadeira média μ

$$
\begin{aligned}
IC(\mu, 0.95) &= \left[\bar{x} - t_{24,0.025}\frac{s}{\sqrt{n}};\ \bar{x} + t_{24,0.025}\frac{s}{\sqrt{n}}\right] \\
&= \left[485 - 2.06\frac{7.1}{5}; 485 + 2.06\frac{7.1}{5}\right] \\
&= [485 - 2.93; 485 + 2.93] \\
&= [482.07; 487.93]
\end{aligned}
$$

Exemplo: Quantas horas de TV por dia?

O histograma a seguir apresenta a distribuição do número de horas de TV assistidas por dia entre os participantes de um estudo em que se coletou uma amostra aleatória.



Exemplo: horas de TV por dia

Encontre um IC de 95% para a média de horas que uma pessoa assiste por dia.

- $n = 905$ pessoas responderam.
- x_i é o número de horas de TV que a pessoa i da amostra assiste.
- $\bar{x} = 1.52$ e $s = 1$
- Erro padrão da média amostral: $s/\sqrt{n} = 0.03$

Exemplo: horas de TV por dia

Utilizamos a distribuição Normal e não a distribuição t, pois n é grande.

- Pelo TCL: $\bar{X}_n \sim N(\mu, \sigma^2/n)$
- O IC de 95% para μ é dado por:

$$
\begin{aligned}
IC(\mu, 0.95) &= \left[\bar{x} - 1.96\frac{s}{\sqrt{n}}\,;\; \bar{x} + 1.96\frac{s}{\sqrt{n}}\right] \\
&= [1.52 - 0.06; 1.52 + 0.06] \\
&= [1.46\,;\; 1.58]
\end{aligned}
$$

Com grau de confiança igual a 95%, estimamos que a média populacional de horas de TV está entre 1.46 e 1.58 horas.

Exemplo - Leite materno

O Ministério da Saúde está preocupado com quantidade de um certo componente tóxico no leite materno.

Em uma amostra de 20 mulheres, a quantidade do componente para cada uma foi:

```
## 16 0 0 2 3 6 8 2 5 0 12 10 5 7 2 3 8 17 9 1
```

Obtenha um intervalo de confiança de 95% para a quantidade média do componente no leite materno.

Exemplo - Leite Materno

Média e desvio padrão da amostra ($n = 20$): $\bar{x} = 5.8$ e $s = 5.08$.

Pela distribuição t, para $\alpha = 0.05$: $t_{19,0.025} = 2.093$

Portanto, o IC de 95% é dado por:

$$
\begin{aligned}
IC(\mu, 0.95) &= \left[\bar{x} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}\ ;\ \bar{x} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}\right] \\
&= [5.8 - 2.39; 5.8 + 2.39] \\
&= [3.41\ ;\ 8.19]
\end{aligned}
$$

Com grau de confiança igual a 95%, estimamos que a média da quantidade do componente entre as mulheres está entre 3.41 e 8.19.

Exemplo - Exame

O desvio padrão da pontuação em um certo exame é 11.3. Uma amostra aleatória de 81 estudantes que fizeram o exame foi coletada e a nota de cada estudantes foi anotada. A pontuação média entre os estudantes amostrados foi 74.6.

Encontre um intervalo de 90% de confiança para a pontuação média entre todos os estudantes que fizeram o exame.

$\bar{x} = 74.6$, $\sigma = 11.3$, $n = 81$, $\alpha = 0.10$ e $z_{0.05} = 1.645$

$$
\begin{aligned}
IC(\mu, 0.90) &= \left[74.6 - 1.645\frac{11.3}{9};\ 74.6 + 1.645\frac{11.3}{9}\right] \\
&= [74.6 - 2.07;\ 74.6 + 2.07] = [72.53;\ 76.67]
\end{aligned}
$$

Com grau de confiança igual a 90%, estimamos que a pontuação média entre os estudantes está entre 72.53 e 76.67.

Exemplo: precisão e tamanho amostral

Qual deve ser o tamanho de uma amostra cuja população da qual ela será sorteada possui um desvio padrão igual a 10, para que a diferença da média amostral para a média da população, em valor absoluto, seja menor que 1, com coeficiente de confiança igual a:

- 95%
- 99%

Fonte: Morettin & Bussab, Estatística Básica 5^a edição, pág 308.

Exemplo: precisão e tamanho amostral

Pelo TCL: $\bar{X}_n \sim N(\mu, 10^2/n)$

Queremos $P(|\bar{X} - \mu| < 1) = 0.95$

$$P(-1 < \bar{X} - \mu < 1) = 0.95$$

$$P\left(-\frac{1}{10/\sqrt{n}} < \frac{\bar{X} - \mu}{10/\sqrt{n}} < \frac{1}{10/\sqrt{n}}\right) = 0.95$$

O que é equivalente a

$$P\left(-\sqrt{n}/10 < Z < \sqrt{n}/10\right) = 0.95$$

Como $P(-1.96 < Z < 1.96) = 0.95$, então

$$\frac{\sqrt{n}}{10} = 1.96 \quad \Rightarrow \quad n \approx 385$$

Exemplo: precisão e tamanho amostral

De modo análogo, para um grau de confiança de 99%, temos que

$$P(-2.58 < Z < 2.58) = 0.99$$

Então,

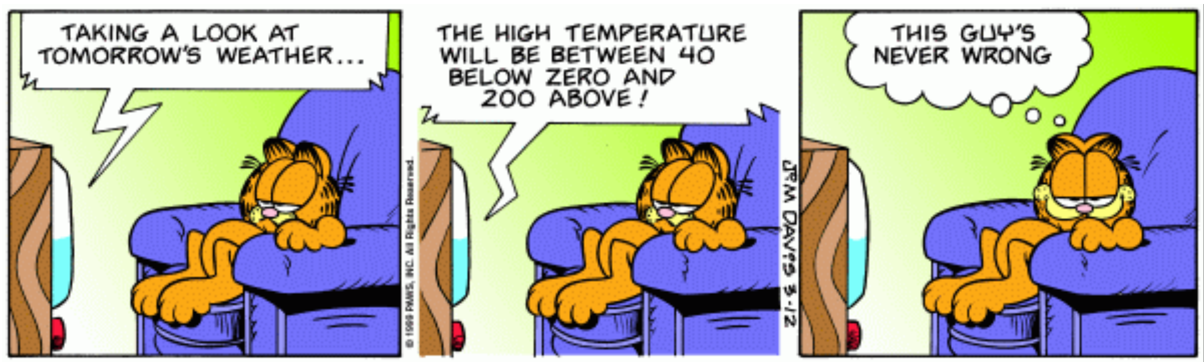
$$\sqrt{n}/10 = 2.58 \quad \Rightarrow \quad n \approx 665$$

Em geral, como já dissemos anteriormente, para uma margem de erro m:

$$n = \left(\frac{z_{\alpha/2}}{m}\right)^2 \sigma^2$$

Leituras

- Ross: capítulo 8.
- OpenIntro: seção 4.2.
- Magalhães: capítulo 7.



Slides produzidos pelos professores:

- Samara Kiihl
- Tatiana Benaglia
- Larissa Matos
- Benilton Carvalho