

Slides Semana 8

Estimadores

Em nossas considerações fizemos a suposição de que o tamanho da amostra n era conhecido e fixo. Na prática, pode ocorrer de estarmos interessado em determinar o tamanho da amostra, de modo que o erro a ser cometido nas inferências seja previamente estipulado, com um determinado grau de confiança. Antes de obtermos as fórmulas para o cálculo do tamanho amostral, devemos conhecer as seguintes definições:

Estimador É uma estatística (característica numérica ou função de dados amostrais) utilizada para estimar um parâmetro desconhecido da população.

Notação:

- T - estimador
- θ - parâmetro.

Exemplos:

1. A média amostral, $T = \bar{X}$, é um estimador para a média populacional $\theta = \mu$.
2. A variância amostral, $T = S^2$, é um estimador para a variância populacional $\theta = \sigma^2$.
3. A proporção amostral, $T = \hat{p}$, é um estimador para a proporção populacional $\theta = p$.

Estimativa É um particular valor numérico que o estimador assume após observada a amostra.

Notação: $T = t$ é uma estimativa do estimador T .

Exemplos:

1. Uma estimativa para a média seria $\bar{X} = \bar{x}$, onde \bar{x} é qualquer valor numérico.
2. Uma estimativa para a variância seria $S^2 = s^2$, onde s^2 é qualquer valor numérico positivo.

Erro Amostral:

Seja θ um parâmetro da distribuição de uma variável aleatória X (parâmetro desconhecido da população) e $T = f(X_1, \dots, X_n)$ um estimador de θ . Então, o erro cometido ao estimar o parâmetro θ pelo estimador T é denominado de **erro amostral** e é definido como:

$$e = T - \theta.$$

Intervalos de confiança

- Vimos que podemos utilizar uma estatística, como $\bar{X} (\hat{p})$, para estimar um parâmetro populacional, como a média populacional μ (proporção populacional p).
- Após coletarmos uma amostra aleatória calculamos \bar{x} , que é a nossa estimativa para μ . Chamamos esta estimativa de **estimativa pontual**.
- Uma estimativa pontual fornece apenas um único valor plausível para o parâmetro. E sabemos que ela pode ser diferente para cada amostra obtida: distribuição amostral.
- O ideal é que se reporte não só a estimativa, mas também a sua imprecisão.
- Duas maneiras: fornecer a estimativa juntamente com o seu **erro padrão** ou fornecer um intervalo de valores plausíveis para o parâmetro de interesse (**intervalo de confiança**).

Suponha que queremos estimar o parâmetro populacional θ através de um intervalo.

Um intervalo de confiança (IC) para θ é sempre da forma:

estimativa \pm margem de erro

$$\hat{\theta} \pm \text{margem de erro}$$

Sendo:

- $\hat{\theta}$ uma estimativa pontual de θ
- **margem de erro:** quantidade que depende da distribuição amostral do estimador pontual de θ , do grau de confiança pré-estabelecido e do erro padrão da estimativa

Intervalo de Confiança como Estimativa de p

Sabemos que pelo Teorema do Limite Central: a distribuição amostral de \hat{p} aproxima-se da seguinte **distribuição Normal** quando n for suficientemente grande:

$$\hat{p} \sim \mathcal{N} \left(p, \frac{p(1-p)}{n} \right)$$

Qual a probabilidade de que o estimador \hat{p} esteja distante do valor verdadeiro, p , em no máximo 1 erro-padrão?

Lembremos que é distante e quem é o erro padrão, i.e.,

$$P \left(|\hat{p} - p| \leq \sqrt{\frac{p(1-p)}{n}} \right).$$

Por outro lado, se padronizamos obtemos:

$$\begin{aligned} P \left(|\hat{p} - p| \leq \sqrt{\frac{p(1-p)}{n}} \right) &= P \left(-\sqrt{\frac{p(1-p)}{n}} \leq \hat{p} - p \leq \sqrt{\frac{p(1-p)}{n}} \right) = P \left(-1 \leq \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq 1 \right) \\ &= P(-1 \leq Z \leq 1) = 0.68 \end{aligned}$$

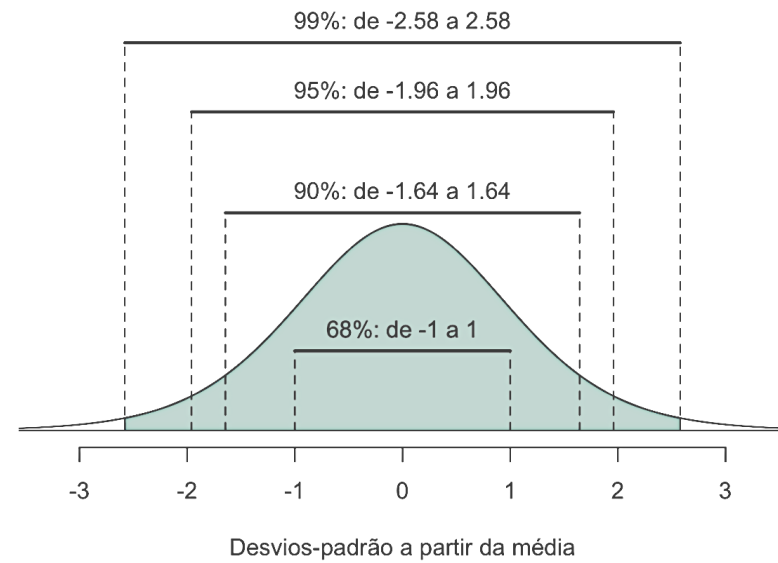
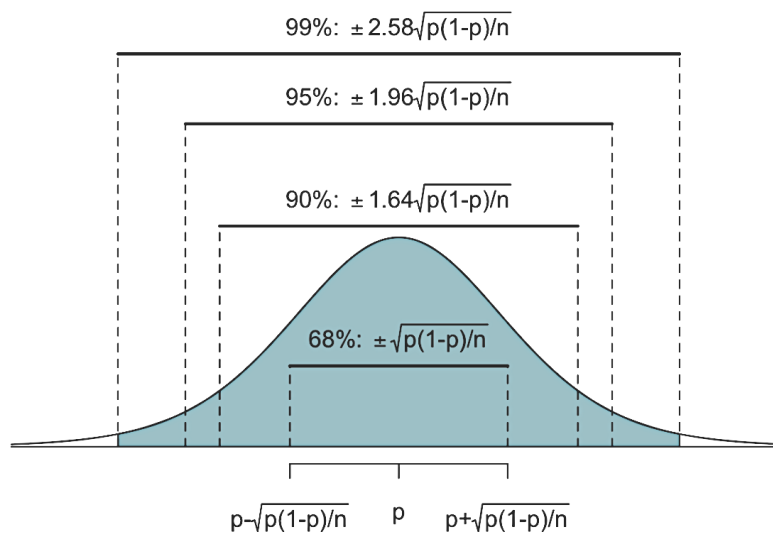
Qual a probabilidade de que o estimador \hat{p} esteja distante do valor verdadeiro, p , em no máximo 1.96 erro-padrão?

$$P \left(|\hat{p} - p| \leq 1.96 \sqrt{\frac{p(1-p)}{n}} \right)$$

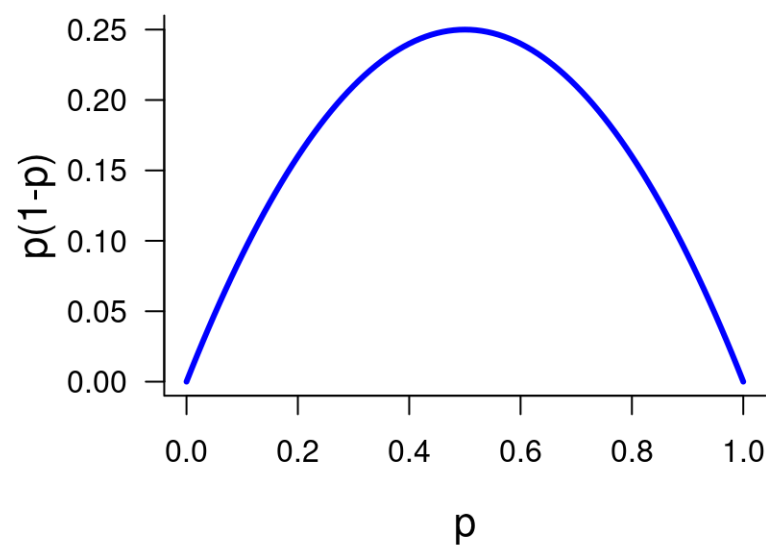
$$\begin{aligned} P \left(|\hat{p} - p| \leq 1.96 \sqrt{\frac{p(1-p)}{n}} \right) &= P \left(-1.96 \sqrt{\frac{p(1-p)}{n}} \leq \hat{p} - p \leq 1.96 \sqrt{\frac{p(1-p)}{n}} \right) \\ &= P \left(-1.96 \leq \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq 1.96 \right) \\ &= P(-1.96 \leq Z \leq 1.96) \\ &= 0.95 \end{aligned}$$

$$\text{Intervalo de confiança de 95\% é } IC(p, 95\%) = \left[\hat{p} - 1.96 \sqrt{\frac{p(1-p)}{n}}; \hat{p} + 1.96 \sqrt{\frac{p(1-p)}{n}} \right]$$

$$\text{Intervalo de confiança de 90\% é } IC(p, 90\%) = \left[\hat{p} - 1.64 \sqrt{\frac{p(1-p)}{n}}; \hat{p} + 1.64 \sqrt{\frac{p(1-p)}{n}} \right]$$



O problema é que em situações práticas não sabemos o valor de $p(1 - p)$, porém:



A função $p(1 - p)$ atinge o valor máximo quando $p = 1/2$, ou seja, $p(1 - p) \leq \frac{1}{4}$.

Intervalo de confiança para p

Vimos que $p(1 - p) \leq \frac{1}{4}$, então erro padrão é maximizado por:

$$\sqrt{\frac{p(1-p)}{n}} \leq \sqrt{\frac{1}{4n}} \iff -\sqrt{\frac{p(1-p)}{n}} \geq -\sqrt{\frac{1}{4n}}$$

Portanto, $IC(p, 95\%) = \left[\hat{p} - 1.96\sqrt{\frac{1}{4n}}; \hat{p} + 1.96\sqrt{\frac{1}{4n}} \right]$.

Caso geral (conservador): Um IC de $100(1 - \alpha)\%$ para p é dado por

$$IC(p, 1 - \alpha) = \left[\hat{p} - z_{\alpha/2} \sqrt{\frac{1}{4n}}; \hat{p} + z_{\alpha/2} \sqrt{\frac{1}{4n}} \right]$$

em que $z_{\alpha/2}$ é tal que:

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

Como encontrar $z_{\alpha/2}$

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

Seja $Z \sim N(0, 1)$. O percentil $z_{\alpha/2}$ é tal que $1 - \alpha = P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2})$

Como determinar $z_{\alpha/2}$?

$$\begin{aligned} 1 - \alpha &= P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = P(Z \leq z_{\alpha/2}) - P(Z \leq -z_{\alpha/2}) \\ &= P(Z \leq z_{\alpha/2}) - P(Z \geq z_{\alpha/2}) \\ &= P(Z \leq z_{\alpha/2}) - [1 - P(Z \leq z_{\alpha/2})] \\ &= 2P(Z \leq z_{\alpha/2}) - 1 \\ &= 2\Phi(z_{\alpha/2}) - 1 \end{aligned}$$

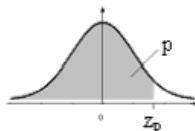
$$\text{Portanto, } 1 - \frac{\alpha}{2} = \Phi(z_{\alpha/2}) \quad \Rightarrow \quad \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = z_{\alpha/2}$$

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 $0.90 = P(-z_{0.05} \leq Z \leq z_{0.05})$.

Tabela I: Distribuição Normal Padrão Acumulada



Fornece $\Phi(z) = P(-\infty < Z \leq z)$, 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	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, $z_{0.05} = 1.64$.

Exemplo

Numa pesquisa de mercado, $n = 400$ pessoas foram entrevistadas (usando amostra aleatória) sobre preferência do produto da marca A, e 60% destas pessoas preferiam a marca A.

Encontre um IC de 95% para a proporção de pessoas que preferem a marca A.

Pelo resultado da pesquisa, $\hat{p} = 0.6$.

Logo, o IC com grau de confiança $1 - \alpha = 0.95$ é dado por:

$$\begin{aligned} IC(p, 0.95) &= \left[0.6 - 1.96 \frac{1}{\sqrt{1600}}; 0.6 + 1.96 \frac{1}{\sqrt{1600}} \right] \\ &= [0.6 - 0.049; 0.6 + 0.049] \\ &= [0.551; 0.649] \end{aligned}$$

Exemplo

Suponha que em $n = 400$ entrevistados, tivéssemos obtido $k = 80$ respostas de pessoas que preferem a marca A.

Vamos obter um intervalo de confiança para p , com grau de confiança de 90%:

- $\hat{p} = \frac{80}{400} = 0.2$

- $1 - \alpha = 0.90$. Então $\alpha/2 = 0.05 \rightarrow z_{\alpha/2} = z_{0.05} = 1.64$

$$\begin{aligned} IC_1(p, 0.90) &= \left[0.2 - 1.64 \frac{1}{\sqrt{1600}}; 0.2 + 1.64 \frac{1}{\sqrt{1600}} \right] \\ &= [0.2 - 0.041; 0.2 + 0.041] \\ &= [0.159; 0.241] \end{aligned}$$

E se usarmos a estimativa \hat{p} para também estimar o erro padrão $\sqrt{\frac{p(1-p)}{n}}$?

Podemos construir o seguinte IC de $100(1 - \alpha)\%$

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

Para os dados do exemplo anterior,

$$\begin{aligned} IC_2(p, 0.90) &= \left[0.2 - 1.64 \sqrt{\frac{(0.2)(0.8)}{400}}; 0.2 + 1.64 \sqrt{\frac{(0.2)(0.8)}{400}} \right] \\ &= [0.2 - 0.033; 0.2 + 0.033] \\ &= [0.167; 0.233] \end{aligned}$$

Exemplo

O intervalo que utiliza \hat{p} também para estimar o erro padrão tem menor margem de erro e, portanto, menor amplitude do que o intervalo que utiliza o fato de $p(1 - p) \leq \frac{1}{4}$. Por isso esse último é chamado de **conservador**.

Veja as amplitudes dos IC 's que encontramos no exemplo anterior:

$$\cdot IC_1(p, 0.90) = [0.159; 0.241] \Rightarrow A_1 = 0.241 - 0.159 = 0.082$$

$$\cdot IC_2(p, 0.90) = [0.167; 0.233] \Rightarrow A_2 = 0.233 - 0.167 = 0.066$$

A amplitude é o dobro da margem de erro.

Determinando o Tamanho Amostral n

Suponha, por exemplo, que queremos estimar a média populacional, μ , então um estimador natural para este parâmetro é a média amostral, \bar{X} , a qual é baseada numa amostra de tamanho n .

Agora, do estudo sobre distribuições amostrais, sabemos que:

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

Conseqüentemente, a distribuição do erro amostral é

$$e = \bar{X} - \mu \sim N\left(0, \frac{\sigma^2}{n}\right)$$

A idéia, então, é obter o valor de n , de modo que o erro amostral seja no máximo igual a um certo valor pré-fixado ϵ tal que

$$P(|e| \leq \epsilon) \geq \gamma, \quad 0 < \gamma < 1.$$

Ou seja, obter o valor de n tal que o erro máximo a ser cometido na inferência seja ϵ com uma probabilidade mínima de $(\gamma \cdot 100)\%$.

Agora, note que,

$$P(|e| \leq \epsilon) = P(|\bar{X} - \mu| \leq \epsilon) = \gamma$$

De modo que, dado γ fixo, é possível obtermos o ponto $z_{\gamma/2}$ na Tabela da Distribuição Normal Padrão, tal que

$$P(-z_{\gamma/2} < Z < z_{\gamma/2}) = \gamma.$$

Assim, teremos que

$$z_{\gamma/2} = \frac{\frac{\epsilon}{\sigma}}{\sqrt{n}}.$$

De onde concluímos que,

$$n = z_{\gamma/2}^2 \cdot \frac{\sigma^2}{\epsilon^2}.$$

Observações Importantes

- Note que ϵ e $z_{\gamma/2}$ podem ser previamente estabelecidos, mas a variância po-pulacional, σ^2 , pode ser desconhecida. Neste caso, torna-se necessário alguma informação prévia sobre a variabilidade da característica populacional que se pretende estudar.

A informação prévia, pode ser obtida, por exemplo, através de uma amostra piloto (amostra prévia), na qual se estima σ^2 por s^2 . Assim, se após substituir σ^2 por s^2 na fórmula, o tamanho amostral obtido for maior que o tamanho da amostra piloto, deve-se obter uma nova amostra com o novo tamanho amostral.

- No caso em que desejamos estimar a proporção populacional, podemos usar o fato de que

$$\hat{p} = \bar{X} \sim N\left(p, \frac{p(1-p)}{n}\right),$$

de modo que

$$n = z_{\gamma/2}^2 \cdot \frac{p(1-p)}{\epsilon^2}.$$

Mas, na prática, o que se faz é usar a fórmula:

$$n = z_{\gamma/2}^2 \cdot \frac{1}{4\epsilon^2}.$$

Levando-se em consideração que o valor de p é desconhecido e que nenhuma informação prévia existe.

Esta última fórmula baseia-se no fato de que a variância do estimador \hat{p} é no máximo igual a $1/4n$, o que leva ao maior tamanho amostral possível. Matematicamente, verifica-se que

$$p(1 - p) \leq 1/4, \text{ onde } 0 < p < 1.$$

Exemplo: Datafolha

A Datafolha quer fazer uma pesquisa de boca-de-urna para predizer o resultado de uma eleição com apenas dois candidatos.

Seleciona então uma a.a. de eleitores e pergunta em quem cada um votou. Para esta pesquisa, o Datafolha quer uma margem de erro de 4%. Qual o tamanho de amostra necessário?

- O grau de confiança é 95% e $IC(p, 0.95) = \hat{p} \pm 1.96 \times EP(\hat{p})$
- Erro padrão de \hat{p} é $EP(\hat{p}) = \sqrt{p(1 - p)/n}$
- Margem de erro: $1.96 \times EP(\hat{p}) = 1.96\sqrt{p(1 - p)/n}$

- Margem de erro desejada é 0.04. Então, o tamanho amostral necessário n é:

$$1.96\sqrt{\frac{p(1-p)}{n}} = 0.04 \quad \Rightarrow \quad n = \frac{1.96^2 p(1-p)}{0.04^2}$$

O problema é que não conhecemos p .

Assim como para encontrar os IC 's, podemos usar o método conservador ou então usar informações obtidas em pesquisas anteriores (caso existam).

Método Conservador: Lembre que $p(1-p)/n$ é a variância da estimativa \hat{p} e já vimos anteriormente que $p(1-p) \leq 1/4$. Então,

$$n = \frac{1.96^2 \times (1/4)}{0.04^2} = 600$$

Outra alternativa: O Datafolha fez uma pesquisa na semana passada e o resultado foi 58% votariam no candidato A e 42% no B . Podemos usar então estas estimativas:

$$n = \frac{1.96^2 \hat{p}(1-\hat{p})}{0.04^2} = \frac{1.96^2 (0.58)(0.42)}{0.04^2} = 585$$

- Uma a.a. de tamanho 585 deverá resultar numa margem de erro de 4% para um IC de 95% para a proporção da população que vota no candidato A .

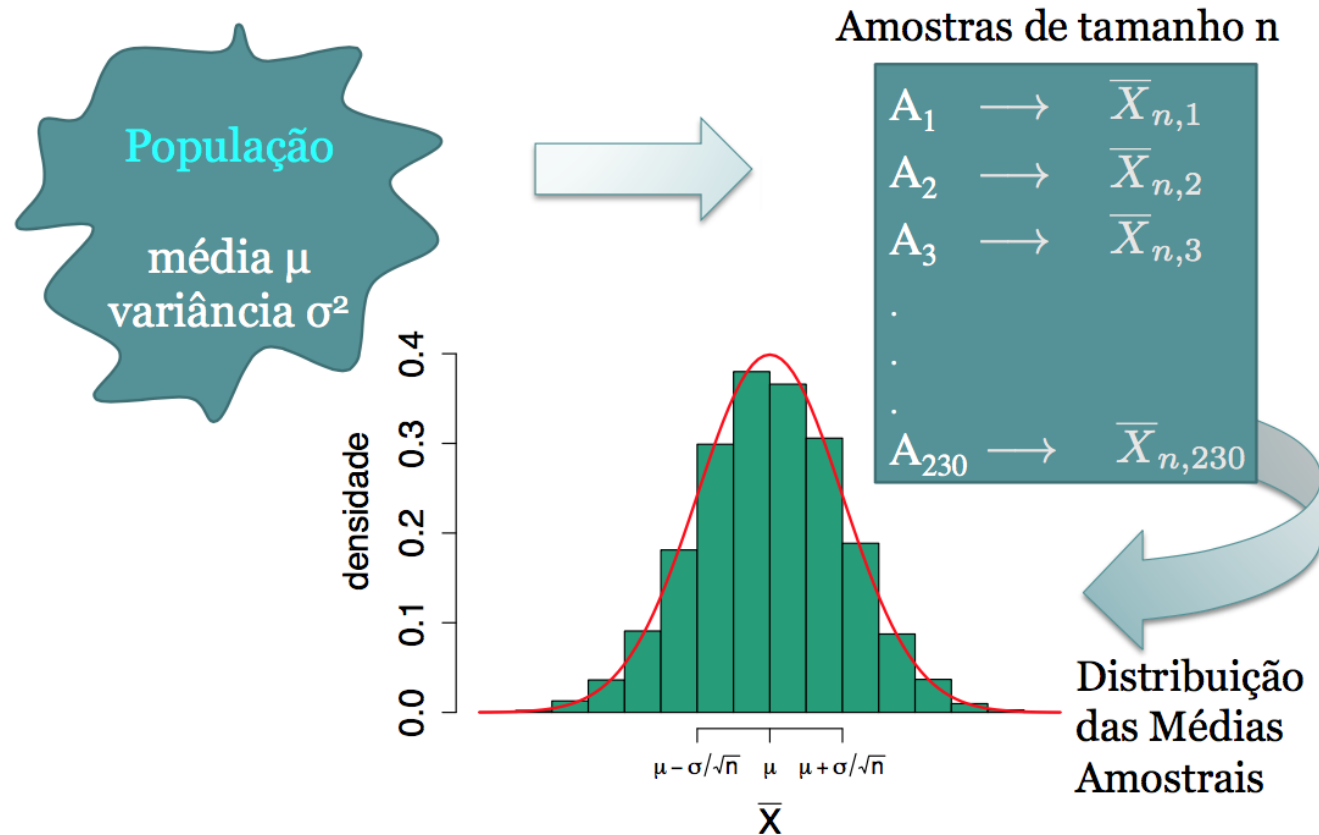
Voltando para a distribuição Amostral de \bar{X}_n

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



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

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

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

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

Intervalo de Confiança para μ : σ conhecido

Seja X_1, \dots, 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 de 95%

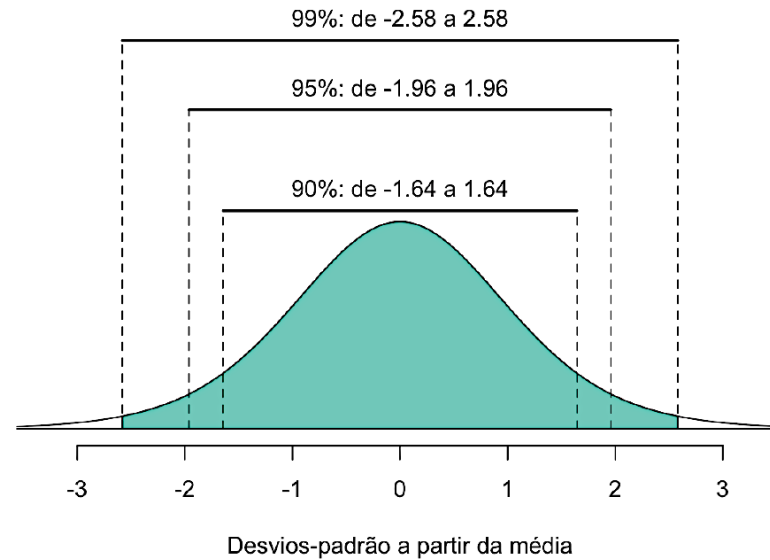
$$IC(\mu, 95\%) = [\bar{x} - 1.96 \sigma/\sqrt{n}; \bar{x} + 1.96 \sigma/\sqrt{n}]$$

Intervalo de confiança de 90%

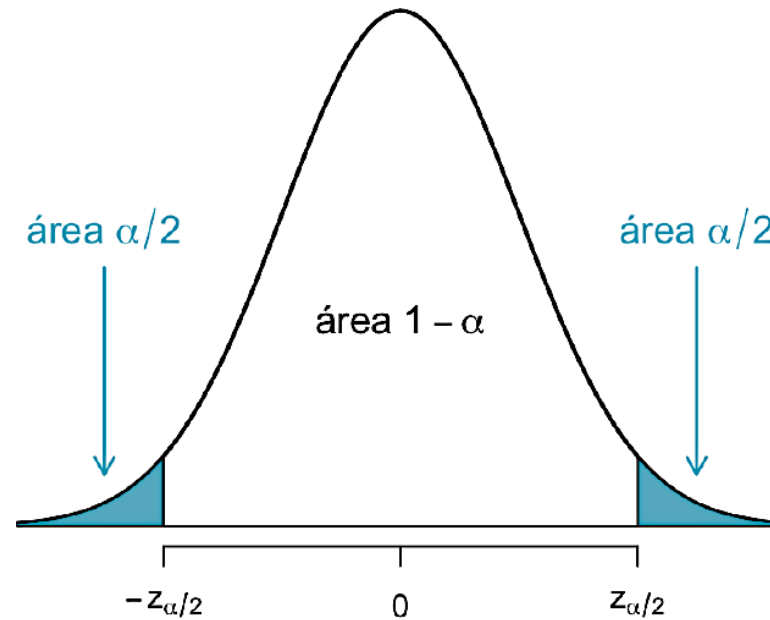
$$IC(\mu, 90\%) = [\bar{x} - 1.64 \sigma/\sqrt{n}; \bar{x} + 1.64 \sigma/\sqrt{n}]$$

Como encontrar $z_{\alpha/2}$

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



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

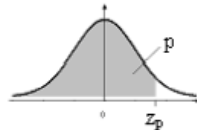


Procure na tabela o valor de z tal que a probabilidade acumulada até o valor de z , isto é $P(Z \leq z) = \Phi(z)$, seja $1 - \alpha/2$.

Exemplo

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

Tabela I: Distribuição Normal Padrão Acumulada



Fornece $\Phi(z) = P(-\infty < Z \leq z)$, 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)

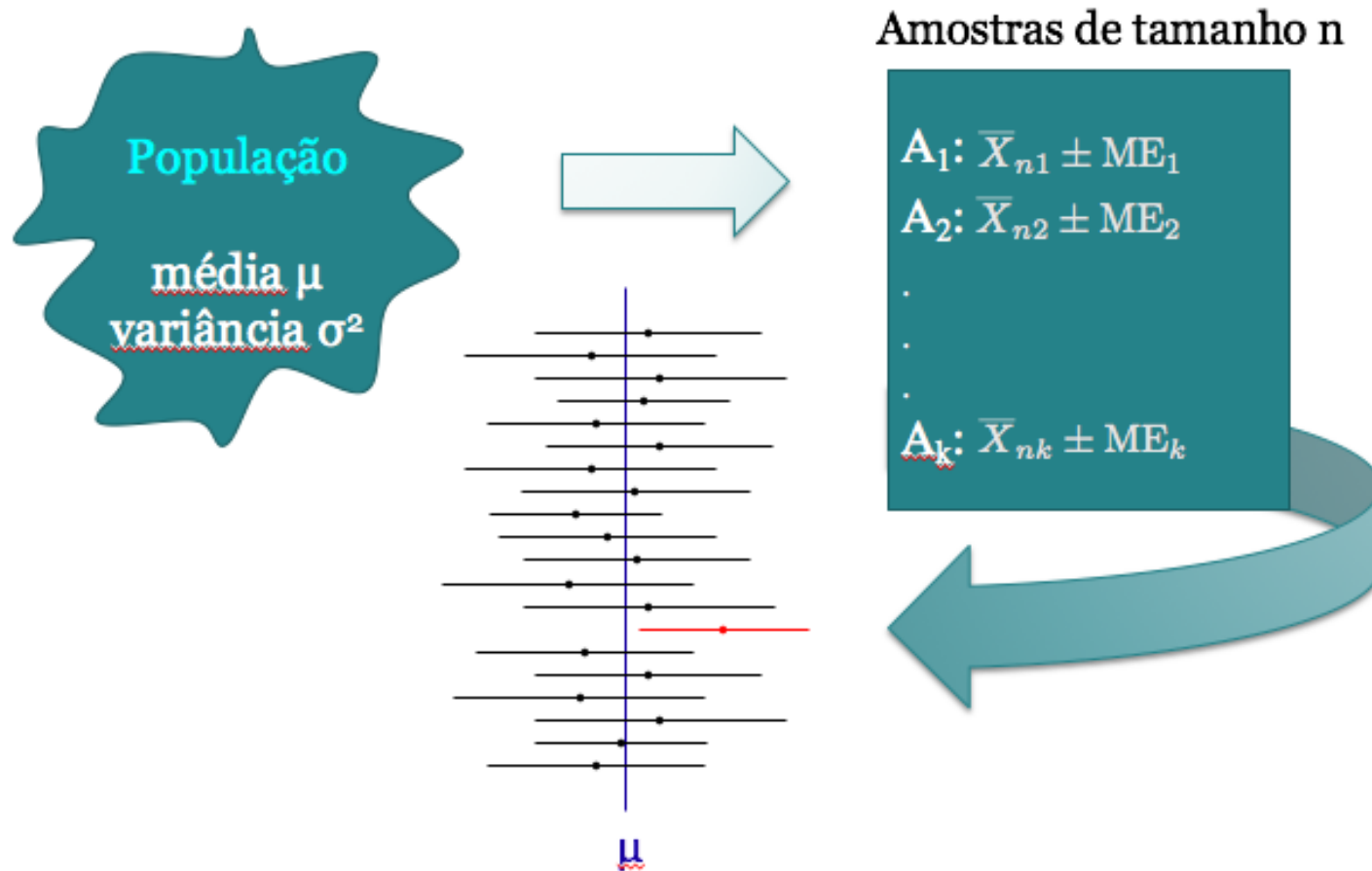
z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	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	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, \dots, 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?

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

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

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

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

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

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, \dots, 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:

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

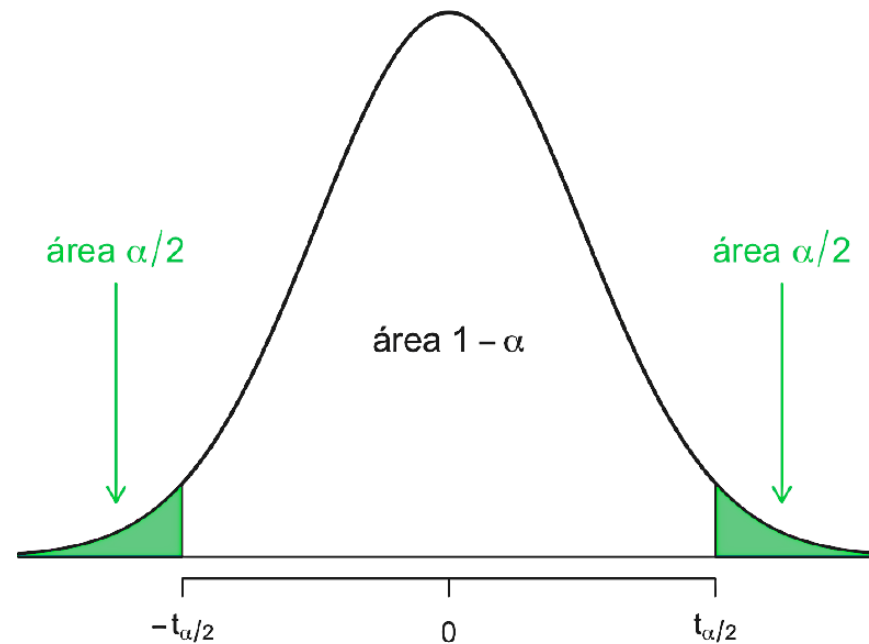
$$P(-t_{n-1,\alpha/2} < T < t_{n-1,\alpha/2}) = 1 - \alpha$$

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

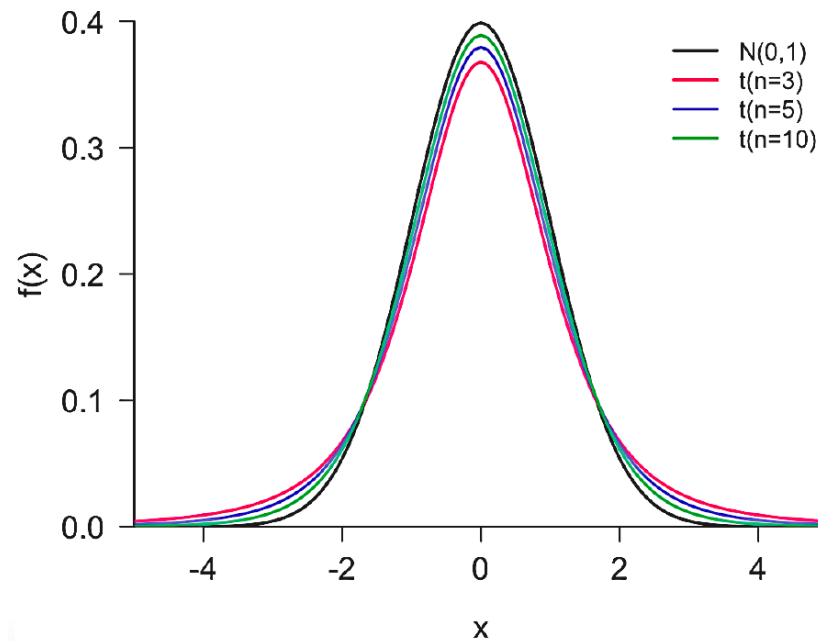
$$P(-t_{n-1,\alpha/2} < T < t_{n-1,\alpha/2}) = 1 - \alpha$$



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**.

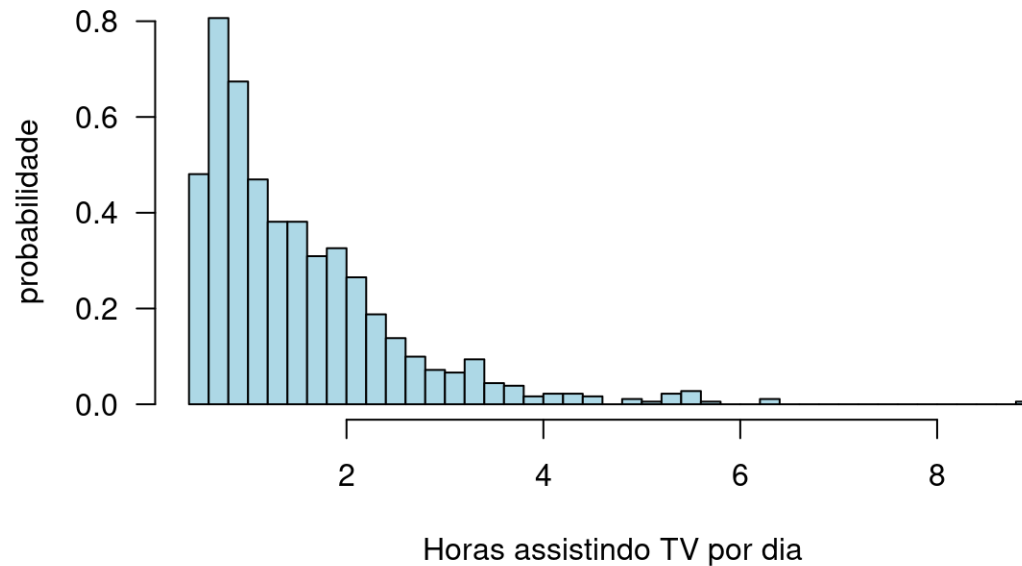
Lembre-se que uma amostra de 25 pacotes apresentou uma média de 485g. Observou-se um desvio padrão na amostra de 7.1g

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: $s/\sqrt{n} = 0.03$

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

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