

# Inferência Estatística II

Prof. Fernando de Souza Bastos  
fernando.bastos@ufv.br

Departamento de Estatística  
Programa de Pós-Graduação em Estatística Aplicada e Biometria  
Universidade Federal de Viçosa  
Campus UFV - Viçosa



# Sumário

- 1 Intervalo de Confiança
- 2 Intervalo de Confiança para a Diferença de Médias
- 3 Intervalo de Confiança para a Diferença de Proporção

Discutiremos inicialmente, dois casos. O primeiro baseado no Teorema Central do Limite. O segundo iremos assumir que a amostra aleatória provém de uma distribuição normal.

Seja  $X_1, \dots, X_n$  uma a.a. de uma distribuição dependendo de um parâmetro  $\theta$ . Seja  $\theta_0$  o valor verdadeiro do parâmetro desconhecido  $\theta$ . Seja  $T$  uma estatística para  $\theta_0$  satisfazendo a seguinte convergência em distribuição

$$\sqrt{n}(T - \theta_0) \xrightarrow{D} N(0, \sigma_T^2) \quad (\star)$$

Em que  $\sigma_T^2$  é a variância assintótica de  $\sqrt{n}T$ . Para nossos propósitos iniciais suponha que  $\sigma_T^2$  é conhecido (geralmente ele é desconhecido)

Seja  $X_1, \dots, X_n$  uma a.a. de uma distribuição dependendo de um parâmetro  $\theta$ . Seja  $\theta_0$  o valor verdadeiro do parâmetro desconhecido  $\theta$ . Seja  $T$  uma estatística para  $\theta_0$  satisfazendo a seguinte convergência em distribuição

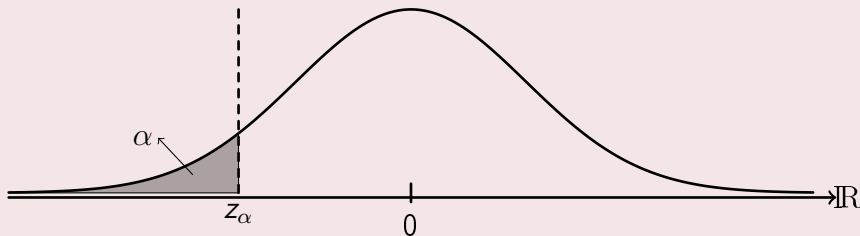
$$\sqrt{n}(T - \theta_0) \xrightarrow{D} N(0, \sigma_T^2) \quad (\star)$$

Em que  $\sigma_T^2$  é a variância assintótica de  $\sqrt{n}T$ . Para nossos propósitos iniciais suponha que  $\sigma_T^2$  é conhecido (geralmente ele é desconhecido)

De  $(\star)$ , temos que

$$\frac{\sqrt{n}(T - \theta_0)}{\sigma_T} \xrightarrow{D} N(0, 1)$$

Seja  $z_\alpha$  o quantil  $\alpha$  da distribuição normal padrão. Ou seja, se  $Z \sim N(0, 1)$ , então  $P(Z < z_\alpha) = \alpha$ .

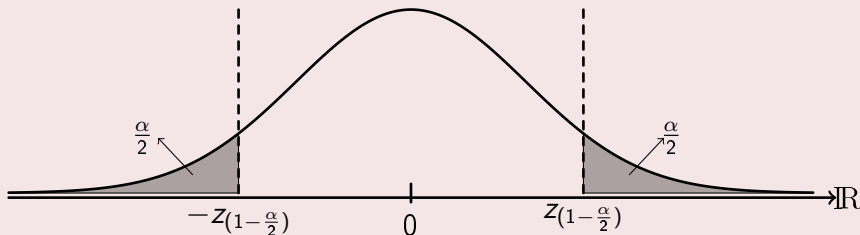


Como

$$\frac{\sqrt{n}(T - \theta_0)}{\sigma_T} \xrightarrow{D} N(0, 1),$$

Temos que

$$P(-z_{1-\frac{\alpha}{2}} < \frac{\sqrt{n}(T - \theta_0)}{\sigma_T} < z_{1-\frac{\alpha}{2}}) \approx 1 - \alpha$$



$$P(-z_{(1-\frac{\alpha}{2})} < \frac{\sqrt{n}(T - \theta_0)}{\sigma_T} < z_{(1-\frac{\alpha}{2})}) \approx 1 - \alpha$$

$$\Rightarrow P(T - \frac{\sigma_T z_{(1-\frac{\alpha}{2})}}{\sqrt{n}} < \theta_0 < T + \frac{\sigma_T z_{(1-\frac{\alpha}{2})}}{\sqrt{n}}) \approx 1 - \alpha$$

Notem que o intervalo

$$\left(T - \frac{\sigma_T z_{(1-\frac{\alpha}{2})}}{\sqrt{n}}; T + \frac{\sigma_T z_{(1-\frac{\alpha}{2})}}{\sqrt{n}}\right)$$

depende de  $T$  e, portanto, é um intervalo aleatório. Com isso, temos que o intervalo aleatório acima contém o valor  $\theta_0$  com probabilidade  $1 - \alpha$  aproximadamente.



Seja  $t$  o valor observado de  $T$ . Então, o intervalo

$$\left( t - \frac{\sigma_T Z(1-\frac{\alpha}{2})}{\sqrt{n}}; t + \frac{\sigma_T Z(1-\frac{\alpha}{2})}{\sqrt{n}} \right) \quad (**)$$

contém ou não o valor de  $\theta_0$ . Podemos pensar nisso como um experimento Bernoulli com probabilidade de sucesso  $p$ , sendo a probabilidade do intervalo conter o valor verdadeiro.

Seja  $t$  o valor observado de  $T$ . Então, o intervalo

$$\left( t - \frac{\sigma_T Z(1-\frac{\alpha}{2})}{\sqrt{n}}; t + \frac{\sigma_T Z(1-\frac{\alpha}{2})}{\sqrt{n}} \right) \quad (**)$$

contém ou não o valor de  $\theta_0$ . Podemos pensar nisso como um experimento Bernoulli com probabilidade de sucesso  $p$ , sendo a probabilidade do intervalo conter o valor verdadeiro.

Em geral, a probabilidade de sucesso é, aproximadamente,  $1 - \alpha$ . O intervalo  $(**)$  é chamado de intervalo de confiança para  $\theta_0$ . E,

- $(1 - \alpha)\%$  é a confiança;
- $\alpha$  é conhecido como nível de significância.

De acordo com o slide anterior, podemos escrever:

$$Li(t) = t - \frac{\sigma \tau Z_{(1-\frac{\alpha}{2})}}{\sqrt{n}}$$

$$Ls(t) = t + \frac{\sigma \tau Z_{(1-\frac{\alpha}{2})}}{\sqrt{n}}$$

$$P_{\theta_0} \left( Li(t) \leq \theta_0 \leq Ls(t) \right) = \begin{cases} 0, & \text{se } \theta_0 \notin \left( Li(t); Ls(t) \right) \\ 1, & \text{se } \theta_0 \in \left( Li(t); Ls(t) \right) \end{cases}$$

Na prática não conhecemos  $\sigma_T$ . Seja  $S_T$  um estimador consistente para  $\sigma_T$ . Então, temos que

$$\frac{\sqrt{n}(T - \theta_0)}{S_T} = \left( \frac{\sigma_T}{S_T} \right) \frac{\sqrt{n}(T - \theta_0)}{\sigma_T} \xrightarrow{D} N(0, 1).$$

$\uparrow$   
 $\xrightarrow{P} 1$

$\uparrow$   
 $\xrightarrow{D} N(0, 1)$

$\uparrow$   
 Teorema de Slutsky

Na prática não conhecemos  $\sigma_T$ . Seja  $S_T$  um estimador consistente para  $\sigma_T$ . Então, temos que

$$\frac{\sqrt{n}(T - \theta_0)}{S_T} = \left( \frac{\sigma_T}{S_T} \right) \frac{\sqrt{n}(T - \theta_0)}{\sigma_T} \xrightarrow{D} N(0, 1).$$

$\uparrow$   
 $\xrightarrow{P} 1$

$\uparrow$   
 $\xrightarrow{D} N(0, 1)$

$\uparrow$   
 Teorema de Slutsky

Com isso, obtemos que o intervalo

$$\left( T - \frac{S_T z_{(1-\frac{\alpha}{2})}}{\sqrt{n}}; T + \frac{S_T z_{(1-\frac{\alpha}{2})}}{\sqrt{n}} \right)$$

conterá  $\theta_0$  com probabilidade  $1 - \alpha$ , aproximadamente.

Se  $t$  e  $s_t$  são os valores observados de  $T$  e  $S_T$ , respectivamente, então

$$\left( t - \frac{s_T Z(1-\frac{\alpha}{2})}{\sqrt{n}}; t + \frac{s_T Z(1-\frac{\alpha}{2})}{\sqrt{n}} \right)$$

é um intervalo de confiança  $(1 - \alpha)\%$  aproximadamente para  $\theta_0$ . Em que  $\frac{s_T}{\sqrt{n}}$  é chamado de erro padrão de  $T$ .

# Exemplo: Intervalo de Confiança para Média

Seja  $X_1, \dots, X_n$  uma amostra aleatória de uma distribuição com média  $\mu$  e variância  $\sigma^2$  (ambos conhecidos). Seja  $\bar{X}$  e  $S^2$  a média amostral e a variância amostral, respectivamente. Pelo TCL,

$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \xrightarrow{D} N(0, 1).$$

Além disso,  $S^2$  é um estimador consistente para  $\sigma^2$ , então  $S$  é um estimador consistente para  $\sigma$ . Pelo teorema de Slutsky,

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S} \xrightarrow{D} N(0, 1)$$

Então,

$$\left( \bar{x} - \frac{1,96s}{\sqrt{n}}; \bar{x} + \frac{1,96s}{\sqrt{n}} \right)$$

é um intervalo de confiança de 95%, aproximadamente, para a média  $\mu$ , em que  $\bar{x}$  e  $s$  são os valores observados de  $\bar{X}$  e  $S$ .



## Exemplo: Intervalo de Confiança para $p$

Seja  $X_1, \dots, X_n$  uma amostra aleatória de uma distribuição Bernoulli com parâmetro de sucesso  $p \in (0, 1)$ . Seja  $\hat{p} = \sum_{i=1}^n \frac{X_i}{n}$  a proporção de sucessos, considerando  $P(X_i = 1) = p$  e  $P(X_i = 0) = 1 - p$ . Pelo TCL,

$$\sqrt{n}(\hat{p} - p) \xrightarrow{D} N(0, p(1 - p)),$$

ou, equivalentemente,

$$\frac{\sqrt{n}(\hat{p} - p)}{\sqrt{p(1 - p)}} \xrightarrow{D} N(0, 1).$$

Como  $\hat{p}$  é consistente para  $p$ , temos que

$$\frac{\sqrt{n}(\hat{p} - p)}{\sqrt{\hat{p}(1 - \hat{p})}} \xrightarrow{D} N(0, 1).$$

Com isso,

$$\left( \hat{p} - \frac{z_{(1-\frac{\alpha}{2})} \sqrt{\hat{p}(1 - \hat{p})}}{\sqrt{n}}; \hat{p} + \frac{z_{(1-\frac{\alpha}{2})} \sqrt{\hat{p}(1 - \hat{p})}}{\sqrt{n}} \right)$$

é um intervalo de confiança  $(1 - \alpha)\%$  assintótico.

Outra forma de fazer é utilizar o método delta para encontrar uma função  $g$  tal que  $p(1 - p)[g'(p)]^2 = k$  (constante). Segue que,

$$g'(p) = \frac{\sqrt{k}}{\sqrt{p(1 - p)}}, \text{ fazendo } k = 1, \text{ temos,}$$

$$g'(p) = \frac{1}{\sqrt{p(1 - p)}}$$

# Para

Um intervalo de confiança  $(1 - \alpha)\%$  assintótico para  $p$  pode ser obtido utilizando a seguinte convergência em distribuição:

$$2\sqrt{n}(\arcsen\sqrt{\hat{p}} - \arcsen\sqrt{p}) \xrightarrow{D} N(0, 1)$$

Escreva o intervalo de confiança explicitamente!

# Exemplo: Intervalo de Confiança para $\mu$ sob Normalidade

Seja  $X_1, \dots, X_n$  uma amostra aleatória da distribuição  $N(\mu, \sigma^2)$ , com  $\mu$  e  $\sigma^2$  desconhecidos. Seja  $\bar{X}$  e  $S^2$  a média e a variância amostral, respectivamente. Então,

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

tem distribuição  $t$  de student com  $n-1$  graus de liberdade, ver teorema 3.6.1 do livro do Hogg, página 214 e 215 da oitava edição.

Seja  $t_{\alpha, n-1}$  o quantil  $\alpha$  de uma distribuição  $t$  de student com  $n-1$  graus de liberdade, ou seja,  $P(T < t_{\alpha, n-1}) = \alpha$ , em que  $T \sim t\text{-student}(n-1)$ . Então,

$$P\left(\bar{x} - t_{(1-\frac{\alpha}{2})} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{(1-\frac{\alpha}{2})} \frac{s}{\sqrt{n}}\right) = 1 - \alpha$$

# Para

- Exercícios da seção 4.2: 7, 8, 9, 15, 18, 21.

# Intervalo de Confiança para a Diferença de Médias

Seja  $X_1, \dots, X_{n_1}$  e  $Y_1, \dots, Y_{n_2}$  duas amostras aleatórias da distribuição de  $X$  e  $Y$ , respectivamente, em que  $X$  e  $Y$  são v.a.'s com  $E(X) = \mu_1$ ,  $Var(X) = \sigma_1^2$ ,  $E(Y) = \mu_2$  e  $Var(Y) = \sigma_2^2$ . Suponha independência das amostras. Estamos interessados em construir um intervalo de confiança para  $\Delta = \mu_1 - \mu_2$ .



Seja  $\bar{X} = \frac{\sum_{i=1}^{n_1} X_i}{n_1}$  e  $\bar{Y} = \frac{\sum_{i=1}^{n_2} Y_i}{n_2}$ . Então,  $\bar{\Delta} = \bar{X} - \bar{Y}$  é um estimador não viesado para  $\Delta$ . Seja  $n = n_1 + n_2$ , assumamos que  $\frac{n_1}{n} \rightarrow \lambda_1 > 0$  e que  $\frac{n_2}{n} \rightarrow \lambda_2 > 0$ , com  $\lambda_1 + \lambda_2 = 1$ . Pelo Teorema Central do Limite,

$$\sqrt{n_1}(\bar{X} - \mu_1) \xrightarrow{D} N(0, \sigma_1^2).$$

Segue que,

$$\sqrt{n}(\bar{X} - \mu_1) = \sqrt{\frac{n}{n_1}} \sqrt{n_1}(\bar{X} - \mu_1) \xrightarrow{D} N(0, \frac{\sigma_1^2}{\lambda_1}).$$

Seja  $\bar{X} = \frac{\sum_{i=1}^{n_1} X_i}{n_1}$  e  $\bar{Y} = \frac{\sum_{i=1}^{n_2} Y_i}{n_2}$ . Então,  $\bar{\Delta} = \bar{X} - \bar{Y}$  é um estimador não viesado para  $\Delta$ . Seja  $n = n_1 + n_2$ , assumamos que  $\frac{n_1}{n} \rightarrow \lambda_1 > 0$  e que  $\frac{n_2}{n} \rightarrow \lambda_2 > 0$ , com  $\lambda_1 + \lambda_2 = 1$ . Pelo Teorema Central do Limite,

$$\sqrt{n_1}(\bar{X} - \mu_1) \xrightarrow{D} N(0, \sigma_1^2).$$

Segue que,

$$\sqrt{n}(\bar{X} - \mu_1) = \sqrt{\frac{n}{n_1}} \sqrt{n_1}(\bar{X} - \mu_1) \xrightarrow{D} N(0, \frac{\sigma_1^2}{\lambda_1}).$$

Usamos o fato de  $\sqrt{\frac{n}{n_1}} \xrightarrow{n \rightarrow \infty} \sqrt{\frac{1}{\lambda_1}}$  e  $\sqrt{n_1}(\bar{X} - \mu_1) \xrightarrow{D} N(0, \sigma_1^2)$ .

Da mesma forma,

$$\sqrt{n}(\bar{Y} - \mu_2) \xrightarrow{D} N(0, \frac{\sigma_2^2}{\lambda_2}).$$

Note que,

$$\begin{aligned}\sqrt{n}[\bar{X} - \bar{Y} - (\mu_1 - \mu_2)] &= \sqrt{n}(\bar{X} - \mu_1) - \sqrt{n}(\bar{Y} - \mu_2) \\ &\xrightarrow{D} N(0, \frac{\sigma_1^2}{\lambda_1} + \frac{\sigma_2^2}{\lambda_2})\end{aligned}$$

Logo,

$$\frac{\sqrt{n}(\bar{X} - \bar{Y} - \Delta)}{\sqrt{\frac{\sigma_1^2}{\lambda_1} + \frac{\sigma_2^2}{\lambda_2}}} \xrightarrow{D} N(0, 1)$$

$$\frac{\sqrt{n}(\bar{X} - \bar{Y} - \Delta)}{\sqrt{\frac{\sigma_1^2}{\lambda_1} + \frac{\sigma_2^2}{\lambda_2}}} \xrightarrow{D} N(0, 1)$$

$$\Rightarrow \frac{\sqrt{n}(\bar{X} - \bar{Y} - \Delta)}{\sqrt{\frac{\sigma_1^2}{\frac{n_1}{n}} + \frac{\sigma_2^2}{\frac{n_2}{n}}}} = \frac{(\bar{X} - \bar{Y} - \Delta)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \xrightarrow{D} N(0, 1)$$

Substituindo  $\sigma_1^2$  e  $\sigma_2^2$  por  $S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^n (X_i - \bar{X})$  e  $S_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^n (Y_i - \bar{Y})$ , respectivamente, temos que,

$$\frac{(\bar{X} - \bar{Y} - \Delta)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \xrightarrow{D} N(0, 1)$$

pois  $S_1^2$  e  $S_2^2$  são consistentes para  $\sigma_1^2$  e  $\sigma_2^2$ , respectivamente.

Logo, temos que um intervalo de confiança  $(1 - \alpha)100\%$ , assintótico para  $\Delta$  fica dado por,

$$\left( \bar{x} - \bar{y} - z_{(1-\frac{\alpha}{2})} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}; \bar{x} - \bar{y} + z_{(1-\frac{\alpha}{2})} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$

Assuma as mesmas condições anteriores. Adicionalmente, suponha que  $X_i \sim \text{Bernoulli}(p_1)$ ,  $i = 1, \dots, n_1$  e  $Y_j \sim \text{Bernoulli}(p_2)$ ,  $j = 1, \dots, n_2$ .

Defina  $\hat{p}_1 = \frac{\sum X_i}{n_1}$  e  $\hat{p}_2 = \frac{\sum Y_j}{n_2}$ . Segue que,

$$\frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \xrightarrow{D} N(0, 1)$$

$$\begin{array}{c} \longleftrightarrow \\ \uparrow \end{array} \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \xrightarrow{D} N(0, 1)$$

Teorema de Slutsky

Com isso, um intervalo de confiança  $(1 - \alpha)100\%$ , assintótico para  $p_1 - p_2$  fica dado por,

$$\left( \hat{p}_1 - \hat{p}_2 \mp z_{(1-\frac{\alpha}{2})} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} \right)$$



# Para

- Exercícios da seção 4.2: 25 ao 27.

# Referências I



Hogg, RV, J McKean e AT Craig (2019). *Introduction to Mathematical Statistics*.