

# 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 Introdução
- 2 Teste da Razão de Verossimilhança
- 3 Exemplo 1 - Hipóteses Bilaterais
- 4 Exemplo 2: Distribuição Exponencial
- 5 Teorema do Teste da Razão de Verossimilhança
- 6 Teste do Tipo Wald
- 7 Teste do Tipo Escore
- 8 Exemplos TRV, Wald e Escore
- 9 Conexão entre os Três Testes

Considere  $X_1, \dots, X_n$  variáveis aleatórias independentes e identicamente distribuídas com função de densidade de probabilidade  $f(x; \theta)$  para  $\theta \in \Omega$ . Considere, ainda, as hipóteses bilaterais:

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0$$

em que  $\theta_0$  é um valor especificado. Lembre-se de que a função de verossimilhança e seu logaritmo são dados por:

$$L(\theta) = \prod_{i=1} f(X_i; \theta) \text{ e } \ell(\theta) = \sum_{i=1} \log f(X_i; \theta)$$

Com  $\hat{\theta}$  sendo a estimativa de máxima verossimilhança de  $\theta$ .

Sabemos, que se  $\theta_0$  é o valor verdadeiro de  $\theta$ , então, assintoticamente,  $L(\theta_0)$  é o valor máximo de  $L(\theta)$ . Considere a razão de duas funções de verossimilhança, a saber,

$$\Lambda = \frac{L(\theta_0)}{L(\hat{\theta})}, \quad \hat{\theta} \in \Theta. \quad (1)$$

Sabemos, que se  $\theta_0$  é o valor verdadeiro de  $\theta$ , então, assintoticamente,  $L(\theta_0)$  é o valor máximo de  $L(\theta)$ . Considere a razão de duas funções de verossimilhança, a saber,

$$\Lambda = \frac{L(\theta_0)}{L(\hat{\theta})}, \quad \hat{\theta} \in \Theta. \quad (1)$$

Observe que  $\Lambda \leq 1$ , pois  $L(\theta_0)$  é uma verossimilhança restrita a  $\Theta_0$  e  $L(\hat{\theta})$  é uma verossimilhança irrestrita a  $\Theta$ . Assim, se  $\theta_0$  é o verdadeiro valor do parâmetro,  $L(\hat{\theta})$  atingirá  $L(\theta_0)$  e  $\Lambda = 1$ , caso contrário,  $L(\theta_0) < L(\hat{\theta})$  e  $\Lambda < 1$ .

Para um nível de significância especificado  $\alpha$ , isso leva à regra de decisão intuitiva:

Rejeitar  $H_0$  em favor de  $H_1$  se  $\Lambda \leq c$ ,

em que  $c$  é tal que  $\alpha = P_{\theta_0}(\Lambda \leq c)$ . Chamamos isso de teste da razão de verossimilhança (TRV).

# Exemplo 1 - Hipóteses Bilaterais

Considere uma amostra aleatória  $X_1, X_2, X_3$  de uma distribuição normal  $N(\theta, 1)$ . Desejamos testar as hipóteses:

$$H_0 : \theta = 5 \quad \text{versus} \quad H_1 : \theta \neq 5,$$

e os dados observados  $X_1 = 6, X_2 = 5, X_3 = 7$ .

Para  $n = 3$ , a função de verossimilhança conjunta é:

$$L(\theta) = \prod_{i=1}^3 \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{(X_i - \theta)^2}{2} \right).$$

- Sob  $H_0$ , temos  $\theta = \theta_0 = 5$ .
- Sob  $H_1$ , o estimador de máxima verossimilhança é

$$\hat{\theta} = \bar{X} = \frac{1}{3}(6 + 5 + 7) = 6$$

.

Substituímos  $\theta_0$  e  $\hat{\theta}$  na função de verossimilhança:

$$\Lambda = \frac{L(\theta_0)}{L(\hat{\theta})} \Rightarrow -2 \ln \Lambda = 2 \sum_{i=1}^3 \frac{(X_i - \theta_0)^2 - (X_i - \hat{\theta})^2}{2}.$$

Ao considerar os dados:

$$\Lambda = e^{-\frac{3}{2}} \Rightarrow -2 \ln \Lambda = 3.$$

**Regra de Decisão:** Sob  $H_0$ , a estatística  $-2 \ln \Lambda$  segue uma distribuição  $\chi^2(1)$ . Para um nível de significância  $\alpha = 0,05$ , o valor crítico é:

$$\chi_{0,05}^2(1) = 3,841.$$

Como  $-2 \ln \Lambda = 3 < 3,841$ , não rejeitamos  $H_0$ .

Com os dados  $X_1 = 6$ ,  $X_2 = 5$ ,  $X_3 = 7$ , não há evidência suficiente para rejeitar  $H_0$  ao nível de significância de 5%.

## Exemplo 2 - Hipóteses Unilaterais

Considere uma amostra aleatória  $X_1, X_2, X_3$  de uma distribuição normal  $N(\theta, 1)$ . Desejamos testar as hipóteses:

$$H_0 : \theta \leq 5 \quad \text{versus} \quad H_1 : \theta > 5,$$

e os dados observados  $X_1 = 6, X_2 = 5.5, X_3 = 6.5$ .

Para  $n = 3$ , a função de verossimilhança conjunta é:

$$L(\theta) = \prod_{i=1}^3 \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{(X_i - \theta)^2}{2} \right).$$

- Sob  $H_0 : \theta \leq \theta_0$ , o valor de  $\theta$  que maximiza a verossimilhança é  $\hat{\theta}_0 = \theta_0 = 5$ .
- Sob  $H_1 : \theta > \theta_0$ , o estimador de máxima verossimilhança é  $\hat{\theta} = \bar{X} = \frac{1}{3}(6 + 5.5 + 6.5) = 6$ .

Substituímos  $\hat{\theta}_0$  e  $\hat{\theta}$  na função de verossimilhança:

$$\Lambda = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})} \Rightarrow -2 \ln \Lambda = 2 \sum_{i=1}^3 \frac{(X_i - \hat{\theta}_0)^2 - (X_i - \hat{\theta})^2}{2}.$$

Para os dados:

$$-2 \ln \Lambda = 3$$

**Regra de Decisão:** Sob  $H_0$ , a estatística  $-2 \ln \Lambda$  segue uma distribuição qui-quadrado com 1 grau de liberdade ( $\chi^2(1)$ ). Para um nível de significância  $\alpha = 0,05$ , o valor crítico é:

$$\chi^2_{0,05}(1) = 3,841.$$

Como  $\bar{X} > \theta_0 = 5$ , e  $-2 \ln \Lambda = 3 < 3,841$ , não rejeitamos  $H_0$ . Com os dados  $X_1 = 6$ ,  $X_2 = 5,5$ ,  $X_3 = 6,5$ , não há evidência suficiente para rejeitar  $H_0$  ao nível de significância de 5%.

# Exemplo 2 - Hipóteses Unilaterais

## Regra de Decisão

Sob  $H_0 : \theta \leq 5$ , considere a estatística de teste:

$$-2 \ln \Lambda.$$

Para rejeitar  $H_0$ :

- ❶ Verifique a direção do desvio:
  - Se  $\hat{\theta} \leq \theta_0$ , não rejeite  $H_0$ .
  - Caso contrário, passe para o próximo passo.
- ❷ Compare  $-2 \ln \Lambda$  com o valor crítico:

$$\chi_{0.05}^2(1) = 3.841.$$

- ❸ Rejeite  $H_0$  se  $-2 \ln \Lambda > 3.841$ .

# Exemplo 1 - Hipóteses Bilaterais

Considere as hipóteses:

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0.$$

## Características:

- O teste avalia se o parâmetro  $\theta$  é exatamente igual a  $\theta_0$  ou se difere em qualquer direção (maior ou menor).
- A função de verossimilhança considera desvios em ambas as direções ( $\theta > \theta_0$  ou  $\theta < \theta_0$ ).

## Região Crítica:

- Rejeitamos  $H_0$  se a estatística do teste  $-2 \ln \Lambda$  for maior que o valor crítico, indicando que  $\theta$  está longe de  $\theta_0$  em qualquer direção.

## Exemplo 2 - Hipóteses Unilaterais

Considere as hipóteses:

$$H_0 : \theta \leq \theta_0 \quad \text{versus} \quad H_1 : \theta > \theta_0.$$

### Características:

- O teste avalia se o parâmetro  $\theta$  é maior que  $\theta_0$ , contra a hipótese de que  $\theta$  seja menor ou igual a  $\theta_0$ .
- A função de verossimilhança considera apenas desvios positivos ( $\theta > \theta_0$ ).

### Região Crítica:

- Rejeitamos  $H_0$  se a estatística do teste  $-2 \ln \Lambda$  for maior que o valor crítico, indicando que há evidência de que  $\theta > \theta_0$ .

# Resumo - Diferença na Região Crítica

## Exemplo 1 - Bilateral

- Considera desvios em ambas as direções ( $\theta > \theta_0$  ou  $\theta < \theta_0$ ).
- Mais rigoroso, pois rejeita  $H_0$  desvios significativos em qualquer direção.

## Exemplo 2 - Unilateral

- Considera apenas desvios em uma direção específica ( $\theta > \theta_0$ ).
- Mais sensível a mudanças direcionais, rejeitando  $H_0$  se houver forte evidência de  $\theta > \theta_0$ .

## Exemplo 3 - Hipóteses Bilaterais

Considere uma amostra aleatória  $X_1, X_2, \dots, X_n$  de uma distribuição  $N(\theta, \sigma^2)$ , em que  $-\infty < \theta < \infty$  e  $\sigma^2 > 0$  são conhecidos. Vamos considerar as hipóteses:

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0,$$

em que  $\theta_0$  é especificado. A função de verossimilhança é dada por:

$$L(\theta) = \left( \frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2 \right\} \quad (2)$$

$$= \left( \frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2 \right\} \exp \left\{ -\frac{1}{2\sigma^2} n(\bar{x} - \theta)^2 \right\}, \quad (3)$$

Em  $\Omega = \{\theta : -\infty < \theta < \infty\}$ , o estimador de máxima verossimilhança é  $\hat{\theta} = \bar{x}$ , e, portanto, a razão de verossimilhança é:

$$\begin{aligned}\Lambda &= \frac{L(\theta_0)}{L(\hat{\theta})} \\ &= \exp \left\{ -\frac{1}{2\sigma^2} n(\bar{X} - \theta_0)^2 \right\}.\end{aligned}$$

Em  $\Omega = \{\theta : -\infty < \theta < \infty\}$ , o estimador de máxima verossimilhança é  $\hat{\theta} = \bar{x}$ , e, portanto, a razão de verossimilhança é:

$$\begin{aligned}\Lambda &= \frac{L(\theta_0)}{L(\hat{\theta})} \\ &= \exp \left\{ -\frac{1}{2\sigma^2} n(\bar{X} - \theta_0)^2 \right\}.\end{aligned}$$

Logo,  $\Lambda \leq c$  é equivalente a  $-2 \log \Lambda \geq -2 \log c$ . No entanto,

$$-2 \log \Lambda = \left( \frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} \right)^2,$$

que segue uma distribuição  $\chi^2(1)$  sob  $H_0$ .

Portanto, o teste da razão de verossimilhança com nível de significância  $\alpha$  afirma que rejeitamos  $H_0$  e aceitamos  $H_1$  quando

$$\left( \frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} \right)^2 \geq \chi_{\alpha}^2(1).$$

Portanto, o teste da razão de verossimilhança com nível de significância  $\alpha$  afirma que rejeitamos  $H_0$  e aceitamos  $H_1$  quando

$$\left( \frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} \right)^2 \geq \chi_\alpha^2(1).$$

- O teste baseado na estatística

$$\left( \frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} \right)^2$$

é equivalente ao teste  $z$ -bilateral porque:

- 1 Ele utiliza o quadrado da estatística  $z$ , que naturalmente segue uma distribuição  $\chi^2(1)$  sob  $H_0$ .

Portanto, o teste da razão de verossimilhança com nível de significância  $\alpha$  afirma que rejeitamos  $H_0$  e aceitamos  $H_1$  quando

$$\left( \frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} \right)^2 \geq \chi^2_{\alpha}(1).$$

- O teste baseado na estatística

$$\left( \frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} \right)^2$$

é equivalente ao teste  $z$ -bilateral porque:

- 1 Ele utiliza o quadrado da estatística  $z$ , que naturalmente segue uma distribuição  $\chi^2(1)$  sob  $H_0$ .
- 2 A bilateralidade está embutida no quadrado da estatística, o que elimina a necessidade de dividir o nível de significância  $\alpha$  entre as duas caudas.

- A razão para usarmos  $\chi^2_{\alpha}(1)$  em vez de  $\chi^2_{\alpha/2}(1)$  é:
  - ① A estatística  $-2 \ln \Lambda$  segue uma distribuição qui-quadrado com 1 grau de liberdade, que já reflete a probabilidade acumulada na cauda superior.

- A razão para usarmos  $\chi^2_\alpha(1)$  em vez de  $\chi^2_{\alpha/2}(1)$  é:
  - ① A estatística  $-2 \ln \Lambda$  segue uma distribuição qui-quadrado com 1 grau de liberdade, que já reflete a probabilidade acumulada na cauda superior.
  - ② A bilateralidade está implícita porque a estatística considera apenas o quadrado das diferenças  $((\bar{X} - \theta_0)^2)$ , incorporando desvios para ambos os lados de  $\theta_0$ .

- A razão para usarmos  $\chi^2_\alpha(1)$  em vez de  $\chi^2_{\alpha/2}(1)$  é:
  - ① A estatística  $-2 \ln \Lambda$  segue uma distribuição qui-quadrado com 1 grau de liberdade, que já reflete a probabilidade acumulada na cauda superior.
  - ② A bilateralidade está implícita porque a estatística considera apenas o quadrado das diferenças  $((\bar{X} - \theta_0)^2)$ , incorporando desvios para ambos os lados de  $\theta_0$ .
  - ③ No caso da estatística  $Z^2 \sim \chi^2(1)$ , rejeitamos  $H_0$  quando:

$$Z^2 \geq \chi^2_\alpha(1),$$

o que é equivalente a:

$$|Z| \geq \sqrt{\chi^2_\alpha(1)}.$$

# Teste da Razão de Verossimilhança para a Distribuição Exponencial

Suponha que  $X_1, \dots, X_n$  são variáveis i.i.d com densidade de probabilidade

$$f(x; \theta) = \theta^{-1} e^{-x/\theta}, \quad \text{para } x, \theta > 0$$

As hipóteses são dadas por

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0.$$

A função de verossimilhança pode ser escrita como:

$$L(\theta) = \theta^{-n} e^{-n\bar{X}/\theta}$$

É fácil ver que a estimativa de máxima verossimilhança de  $\theta$  é  $\bar{X}$ .

Após alguma simplificação, a estatística do teste da razão de verossimilhança pode ser escrita como:

$$\Lambda = e^n \left( \frac{\bar{X}}{\theta_0} \right)^n e^{-n\bar{X}/\theta_0}.$$

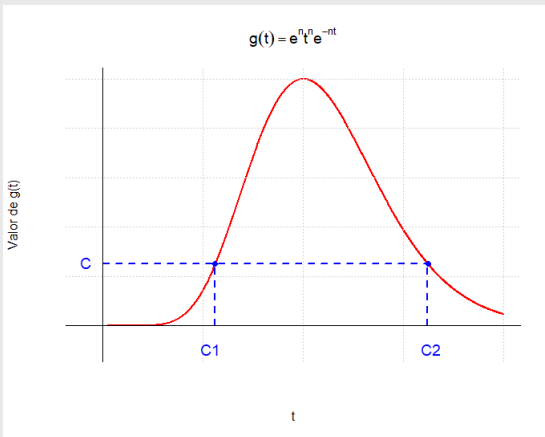
Além da constante  $e^n$ , a estatística do teste tem a forma:

$$g(t) = e^n t^n e^{-nt}, \quad t > 0$$

em que  $t = \bar{X}/\theta_0$ .

Como mostra a Figura abaixo,  $g(t) \leq c$  se, e somente se,  $t \leq c_1$  ou  $t \geq c_2$ . Isso leva a:

$\Lambda \leq c$  se e somente se  $\bar{X}/\theta_0 \leq c_1$  ou  $\bar{X}/\theta_0 \geq c_2$ .



Notem que  $M_{X_i}(t) = (1 - t\theta)^{-1}$ , é a função geradora de momentos de  $X_i$ ,  $i = 1, 2, \dots, n$ . Assim, a função geradora de momentos de  $Y = 2\sum_{i=1}^n X_i/\theta_0$  é dada por  $M_Y(t) = (1 - 2t)^{-2n/2}$ , que é a função geradora de momentos de uma variável  $\chi^2(2n)$ .

Notem que  $M_{X_i}(t) = (1 - t\theta)^{-1}$ , é a função geradora de momentos de  $X_i$ ,  $i = 1, 2, \dots, n$ . Assim, a função geradora de momentos de  $Y = 2\sum_{i=1}^n X_i/\theta_0$  é dada por  $M_Y(t) = (1 - 2t)^{-2n/2}$ , que é a função geradora de momentos de uma variável  $\chi^2(2n)$ .

Ou seja,  $\bar{X}/\theta_0 \leq c_1 \iff Y = 2\sum_{i=1}^n X_i/\theta_0 \leq k_1 = 2nc_1$  ou

Notem que  $M_{X_i}(t) = (1 - t\theta)^{-1}$ , é a função geradora de momentos de  $X_i$ ,  $i = 1, 2, \dots, n$ . Assim, a função geradora de momentos de  $Y = 2\sum_{i=1}^n X_i/\theta_0$  é dada por  $M_Y(t) = (1 - 2t)^{-2n/2}$ , que é a função geradora de momentos de uma variável  $\chi^2(2n)$ .

Ou seja,  $\bar{X}/\theta_0 \leq c_1 \iff Y = 2\sum_{i=1}^n X_i/\theta_0 \leq k_1 = 2nc_1$  ou

$\bar{X}/\theta_0 \geq c_2 \iff Y = 2\sum_{i=1}^n X_i/\theta_0 \geq k_2 = 2nc_2$ .

Baseado nas observações anteriores, podemos utilizar a seguinte regra de decisão para o teste de nível  $\alpha$ :

Baseado nas observações anteriores, podemos utilizar a seguinte regra de decisão para o teste de nível  $\alpha$ :

Rejeitar  $H_0$  se  $(2/\theta_0) \sum_{i=1} X_i \leq \chi^2_{1-\alpha/2}(2n)$  ou  $(2/\theta_0) \sum_{i=1} X_i \geq \chi^2_{\alpha/2}(2n)$ ,

em que  $\chi^2_{1-\alpha/2}(2n)$  é o quantil inferior  $\alpha/2$  de uma distribuição  $\chi^2$  com  $2n$  graus de liberdade e  $\chi^2_{\alpha/2}(2n)$  é o quantil superior  $\alpha/2$  de uma distribuição  $\chi^2$  com  $2n$  graus de liberdade.

# Teorema do Teste da Razão de Verossimilhança

## Teorema 1

*Suponha que as condições de regularidade  $R0$  a  $R5$  são satisfeitas. Sob a hipótese nula,  $H_0 : \theta = \theta_0$ , temos que  $\{-2 \log \Lambda\} \xrightarrow{D} \chi^2(1)$ .*

# Teorema do Teste da Razão de Verossimilhança

## Teorema 1

*Suponha que as condições de regularidade  $R0$  a  $R5$  são satisfeitas. Sob a hipótese nula,  $H0 : \theta = \theta_0$ , temos que  $\{-2 \log \Lambda\} \xrightarrow{D} \chi^2(1)$ .*

## Demonstração:

**Prova:** Expanda a função  $\ell(\theta)$  em um Polinômio de Taylor em torno de  $\theta_0$  de ordem 1 e avalie-a no estimador de máxima verossimilhança,  $\hat{\theta}$ . Isso resulta em

$$\ell(\hat{\theta}) = \ell(\theta_0) + (\hat{\theta} - \theta_0)\ell'(\theta_0) + \frac{1}{2}(\hat{\theta} - \theta_0)^2\ell''(\theta_n^*), \quad (4)$$

em que  $\theta_n^*$  está entre  $\hat{\theta}$  e  $\theta_0$ .

## Demonstração:

Como  $\hat{\theta} \xrightarrow{P} \theta_0$ , segue que  $\theta_n^* \xrightarrow{P} \theta_0$ . Isso, além do fato de que a função  $\ell'(\theta)$  é contínua e a equação (6.2.22) do Teorema 6.2.2 do livro texto (Hogg 8ª edição) implicam que

$$-\frac{1}{n}\ell''(\theta_n^*) \xrightarrow{P} I(\theta_0). \quad (5)$$

Pelo Corolário 6.2.3, do livro texto,

$$\frac{1}{\sqrt{n}}\ell'(\theta_0) = \sqrt{n}(\hat{\theta} - \theta_0)I(\theta_0) + R_n, \quad (6)$$

em que  $R_n \xrightarrow{P} 0$ .

Se substituirmos (5) e (6) na expressão (4) e fizermos algumas simplificações, obtemos

$$-2 \log \Lambda = 2(\ell(\hat{\theta}) - \ell(\theta_0)) = \left[ \sqrt{nI(\theta_0)}(\hat{\theta} - \theta_0) \right]^2 + R_n^*,$$

em que  $R_n^* \xrightarrow{P} 0$ . Pelo Teorema 5.2.4 e pelo Teorema 6.2.2, o primeiro termo no lado direito da equação acima converge em distribuição para uma distribuição  $\chi^2$  com um grau de liberdade. cqd ■

Logo, defina a estatística de teste  $\chi_L^2 = -2 \log \Lambda$ . Para as hipóteses,

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0,$$

este teorema sugere a regra de decisão:

Rejeitar  $H_0$  em favor de  $H_1$  se  $\chi_L^2 \geq \chi_\alpha^2(1)$ .

Pelo último teorema, este teste possui nível assintótico  $\alpha$ . Se não conseguirmos obter a estatística de teste ou sua distribuição em uma forma fechada, podemos usar este teste assintótico.

# Teste do Tipo Wald

Além do teste da razão de verossimilhança, na prática, são empregados outros dois testes relacionados à verossimilhança. Uma estatística de teste natural é baseada na distribuição assintótica de  $\hat{\theta}$ . Considere a estatística

$$\chi_W^2 = \left\{ \sqrt{nI(\hat{\theta})}(\hat{\theta} - \theta_0) \right\}^2.$$

como  $I(\theta)$  é uma função contínua,  $I(\hat{\theta}) \xrightarrow{P} I(\theta_0)$  sob a hipótese nula. Portanto, sob  $H_0$ ,  $\chi_W^2$  possui uma distribuição assintótica  $\chi^2$  com um grau de liberdade.

# Teste do Tipo Wald

Além do teste da razão de verossimilhança, na prática, são empregados outros dois testes relacionados à verossimilhança. Uma estatística de teste natural é baseada na distribuição assintótica de  $\hat{\theta}$ . Considere a estatística

$$\chi_W^2 = \left\{ \sqrt{nI(\hat{\theta})}(\hat{\theta} - \theta_0) \right\}^2.$$

como  $I(\theta)$  é uma função contínua,  $I(\hat{\theta}) \xrightarrow{P} I(\theta_0)$  sob a hipótese nula. Portanto, sob  $H_0$ ,  $\chi_W^2$  possui uma distribuição assintótica  $\chi^2$  com um grau de liberdade.

Isso sugere a regra de decisão

$$\text{Rejeitar } H_0 \text{ em favor de } H_1 \text{ se } \chi_W^2 \geq \chi_\alpha^2(1). \quad (7)$$



O teste (7) é frequentemente referido como um teste do tipo Wald, em homenagem a **Abraham Wald** (nasceu em 31 de outubro de 1902, morreu em 1950 de acidente de avião), que foi um estatístico do século XX. Após sua morte, Wald foi criticado por Sir Ronald A. Fisher. O trabalho de Wald foi defendido, posteriormente, por Jerzy Neyman e outros acadêmicos proeminentes.

# Teste do Tipo Escore

O terceiro teste é chamado de teste de escores de Rao, em homenagem a **Calyampudi Radhakrishna Rao** (10 de setembro de 1920 - 22 de agosto de 2023). A American Statistical Association o descreveu como “uma lenda viva cujo trabalho influenciou não apenas as estatísticas, mas teve implicações de longo alcance para diversos outros campos.” O Times of India listou Rao como um dos 10 maiores cientistas indianos de todos os tempos.



Os escores são os componentes do vetor

$$S(\theta) = \left\{ \frac{\partial \log f(X_1; \theta)}{\partial \theta}, \dots, \frac{\partial \log f(X_n; \theta)}{\partial \theta} \right\}.$$

Em nossa notação, temos

$$\frac{1}{\sqrt{n}} \ell'(\theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial \log f(X_i; \theta_0)}{\partial \theta}.$$

Defina a estatística

$$\chi_R^2 = \left\{ \frac{\ell'(\theta_0)}{\sqrt{nI(\theta_0)}} \right\}^2.$$

Sob  $H_0$ , segue da expressão (6) que

$$\chi_R^2 = \chi_W^2 + R_{0n},$$

em que  $R_{0n}$  converge para 0 em probabilidade. Portanto, a próxima regra de decisão define um teste.

Defina a estatística

$$\chi_R^2 = \left\{ \frac{\ell'(\theta_0)}{\sqrt{nI(\theta_0)}} \right\}^2.$$

Sob  $H_0$ , segue da expressão (6) que

$$\chi_R^2 = \chi_W^2 + R_{0n},$$

em que  $R_{0n}$  converge para 0 em probabilidade. Portanto, a próxima regra de decisão define um teste.

$$\text{Rejeite } H_0 \text{ em favor de } H_1 \text{ se } \chi_R^2 \geq \chi_\alpha^2(1). \quad (8)$$

# Exemplos TRV, Wald e Escore

Seja  $X_1, \dots, X_n$  uma amostra aleatória com uma distribuição  $\text{beta}(\theta, 1)$ . Considere as hipóteses:

$$H_0 : \theta = 1 \text{ versus } H_1 : \theta \neq 1$$

# Exemplos TRV, Wald e Escore

Seja  $X_1, \dots, X_n$  uma amostra aleatória com uma distribuição  $\text{beta}(\theta, 1)$ . Considere as hipóteses:

$$H_0 : \theta = 1 \text{ versus } H_1 : \theta \neq 1$$

A densidade da  $\text{Beta}(\theta, 1)$  é:

$$f(x; \theta) = \theta x^{\theta-1}, \quad 0 < x < 1, \quad \theta > 0.$$

O logaritmo da densidade, que chamaremos de  $\ell(\theta)$ , é:

$$\ell(\theta) = \log f(x; \theta) = \log \theta + (\theta - 1) \log x.$$

É possível mostrar que  $\hat{\theta} = -\frac{n}{\sum_{i=1}^n \log X_i}$  é o estimador de máxima verossimilhança de  $\theta$ . Após alguma simplificação, o valor da função de verossimilhança no estimador de máxima verossimilhança é dado por:

$$L(\hat{\theta}) = \left( -\sum_{i=1}^n \log X_i \right)^{-n} \exp \left\{ n(\log n - 1) - \sum_{i=1}^n \log X_i \right\}.$$

# Exemplos TRV, Wald e Escore

Além disso,  $L(1) = 1$ , portanto, a estatística do teste da razão de verossimilhança é  $\Lambda = \frac{1}{L(\hat{\theta})}$ , de modo que:

$$\chi_L^2 = -2 \log \Lambda = 2 \left( - \sum_{i=1}^n \log X_i - n \log \left( - \sum_{i=1}^n \log X_i \right) - n + n \log n \right)$$

Lembre-se de que a informação para esta distribuição é  $I(\theta) = \theta^{-2}$ .

# Exemplos TRV, Wald e Escore

Para o teste do tipo Wald, podemos estimar isso consistentemente como  $\hat{\theta}^{-2}$ . O teste do tipo Wald simplifica-se para:

$$\chi_W^2 = \sqrt{\frac{n}{\hat{\theta}^2}}(\hat{\theta} - 1) = n \left(1 - \frac{1}{\hat{\theta}}\right)^2$$

# Exemplos TRV, Wald e Escore

Por fim, para o teste do tipo escores,  $\ell'(1)$  é dado por:

$$\ell'(1) = \sum_{i=1}^n \log X_i + n$$

Portanto, a estatística do teste do tipo escores é:

$$\chi_R^2 = \left\{ \frac{\left( \sum_{i=1}^n \log X_i + n \right)}{\sqrt{n}} \right\}^2 .$$

Os testes de Razão de Verossimilhança (LRT), Wald e Escore avaliam a mesma hipótese nula, mas de formas distintas:

- **Razão de Verossimilhança (LRT):** Compara o ajuste de dois modelos, um com e outro sem restrições, por meio da razão das verossimilhanças.

Os testes de Razão de Verossimilhança (LRT), Wald e Escore avaliam a mesma hipótese nula, mas de formas distintas:

- **Razão de Verossimilhança (LRT):** Compara o ajuste de dois modelos, um com e outro sem restrições, por meio da razão das verossimilhanças.
- **Wald:** Avalia diretamente se o parâmetro estimado está próximo do valor hipotético, utilizando a variância do estimador.

Os testes de Razão de Verossimilhança (LRT), Wald e Escore avaliam a mesma hipótese nula, mas de formas distintas:

- **Razão de Verossimilhança (LRT):** Compara o ajuste de dois modelos, um com e outro sem restrições, por meio da razão das verossimilhanças.
- **Wald:** Avalia diretamente se o parâmetro estimado está próximo do valor hipotético, utilizando a variância do estimador.
- **Escore:** Analisa o gradiente da verossimilhança no ponto nulo para identificar evidências de desvio em direção à alternativa.

Os três testes têm como base a função de verossimilhança, mas fazem uso de diferentes aspectos da mesma:

- O **LRT** considera a *diferença máxima* na função de verossimilhança entre os modelos nulo e alternativo.

Os três testes têm como base a função de verossimilhança, mas fazem uso de diferentes aspectos da mesma:

- O **LRT** considera a *diferença máxima* na função de verossimilhança entre os modelos nulo e alternativo.
- O **teste de Wald** utiliza a *distância do estimador* ao valor nulo ponderada pela variância.

Os três testes têm como base a função de verossimilhança, mas fazem uso de diferentes aspectos da mesma:

- O **LRT** considera a *diferença máxima* na função de verossimilhança entre os modelos nulo e alternativo.
- O **teste de Wald** utiliza a *distância do estimador* ao valor nulo ponderada pela variância.
- O **teste de escore** analisa o *declive da verossimilhança* no ponto nulo.

- **Teste de Wald:**

- Fácil de implementar, pois depende apenas do estimador de máxima verossimilhança (EMV) e de sua variância estimada.

- **Teste de Wald:**

- Fácil de implementar, pois depende apenas do estimador de máxima verossimilhança (EMV) e de sua variância estimada.
- Pode ser impreciso em pequenas amostras ou quando o parâmetro está próximo da borda do espaço paramétrico.

- **Teste de Wald:**

- Fácil de implementar, pois depende apenas do estimador de máxima verossimilhança (EMV) e de sua variância estimada.
- Pode ser impreciso em pequenas amostras ou quando o parâmetro está próximo da borda do espaço paramétrico.

- **Teste de Razão de Verossimilhança (LRT):**

- Considerado mais robusto, pois avalia a qualidade do ajuste entre os modelos nulo e alternativo.

- **Teste de Wald:**

- Fácil de implementar, pois depende apenas do estimador de máxima verossimilhança (EMV) e de sua variância estimada.
- Pode ser impreciso em pequenas amostras ou quando o parâmetro está próximo da borda do espaço paramétrico.

- **Teste de Razão de Verossimilhança (LRT):**

- Considerado mais robusto, pois avalia a qualidade do ajuste entre os modelos nulo e alternativo.
- Pode ser computacionalmente caro, já que exige a estimação do modelo sob a hipótese alternativa.

- **Teste de Wald:**

- Fácil de implementar, pois depende apenas do estimador de máxima verossimilhança (EMV) e de sua variância estimada.
- Pode ser impreciso em pequenas amostras ou quando o parâmetro está próximo da borda do espaço paramétrico.

- **Teste de Razão de Verossimilhança (LRT):**

- Considerado mais robusto, pois avalia a qualidade do ajuste entre os modelos nulo e alternativo.
- Pode ser computacionalmente caro, já que exige a estimação do modelo sob a hipótese alternativa.

- **Teste de Escore:**

- Útil quando a estimação do modelo alternativo é difícil, pois depende apenas das derivadas da verossimilhança no ponto nulo.

- **Teste de Wald:**

- Fácil de implementar, pois depende apenas do estimador de máxima verossimilhança (EMV) e de sua variância estimada.
- Pode ser impreciso em pequenas amostras ou quando o parâmetro está próximo da borda do espaço paramétrico.

- **Teste de Razão de Verossimilhança (LRT):**

- Considerado mais robusto, pois avalia a qualidade do ajuste entre os modelos nulo e alternativo.
- Pode ser computacionalmente caro, já que exige a estimação do modelo sob a hipótese alternativa.

- **Teste de Escore:**

- Útil quando a estimação do modelo alternativo é difícil, pois depende apenas das derivadas da verossimilhança no ponto nulo.
- Assume que o modelo nulo está bem especificado, o que pode limitar sua aplicabilidade.

# Para

- **Exercícios da seção 6.3:** 6, 9, 10, 11, 12, 13, 16, 18, 19.

# Referências I



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