

# TESTE DA MEDIANA

UMA PERSPECTIVA VEROSSIMILHANCISTA

Esley Caminhas Ferreira    Renan Carlos da Silva

# Introdução

O teste da mediana é um teste não paramétrico para comparar medianas de dois grupos independentes, útil em casos com quando são violadas suposições de normalidade, ou há presença de *outliers*, ou ainda em pequenas amostras.

Nesta apresentação, nosso objetivo é desenvolver o teste a partir de uma abordagem focada na verossimilhança, modelando a probabilidade de uma observação estar acima da mediana global como um modelo binomial.

# Definição Formal da Verossimilhança

Considere dois grupos independentes e com observações:

- Grupo A:
- Grupo B:  $Y_{A,1}, \dots, Y_{A,n_A}$   
 $Y_{B,1}, \dots, Y_{B,n_B}$

Para cada grupo, definimos:

$$k \in \{A, B\}$$

$$p_k = P(Y_k > \theta)$$

$$\gamma_k = P(Y_k = \theta)$$

onde  $\theta$  é a mediana global dos dados combinados.

$$1 - p_k - \gamma_k = P(Y_k < \theta)$$

# Hipóteses

Hipóteses:

- (As medianas populacionais são iguais)
- 

Isto é, sob temos:

,

onde é o número de observações onde

# Desenvolvendo a verossimilhança

Para cada observação , a contribuição para a verossimilhança é

Para o grupo :

# Desenvolvendo a verossimilhança

onde:

# Verossimilhança Conjunta

**EMV:** Com 4 parâmetros, teríamos certo trabalho para encontrar os estimadores de máxima verossimilhança. Após alguns cálculos, obtemos os estimadores que são intuitivos, isto é:

# EMV

Sob ,



# Teste da Razão de Verossimilhanças

Os graus de liberdade, nesse caso, são 2 porque é a diferença entre os 4 parâmetros que estimamos na verossimilhança irrestrita e os 2 que estimamos sob .

# Simplificando o teste

Sabemos que, como temos em geral dados contínuos (geralmente discretizados, em muitos casos) e, à medida que , o seguinte resultado se aplica:

Isso facilitará nosso procedimento, à medida que agora temos um caso binomial. Isto é: agora sem o termo , a verossimilhança conjunta se reduz a

# Teste da razão de verossimilhanças

Portanto, o teste da razão de verossimilhanças fica:

Sendo que

# Observação

É importante notar que omitir torna o estimador viesado, mas ele ainda é assintoticamente não viesado.

Esta manipulação está em consonância com o teste tradicional, que ignora os valores que são iguais à mediana.

# Exemplo prático

Para um caso prático, vamos usar microdados do ENEM 2023, disponíveis no portal de [dados abertos](https://dadosabertos.gov.br) do gov.com.br.

Neste exemplo, vamos considerar as médias das notas da prova de matemática por município, em 2 grupos de escolas: públicas e particulares.

Consideramos apenas alunos que compareceram a todos os dias de prova e não tiveram nenhum problema com o resultado (como redação anulada ou eliminação da prova, por exemplo). Também desconsideramos as escolas federais.

O objetivo é verificar se a nota média das escolas particulares é superior à das escolas públicas.

# O conjunto de dados

	tipo_escola	municipio	media_matematica
1	Pública	Afonso Cláudio	507,6
2	Pública	Água Doce do Norte	425,1
3	Pública	Águia Branca	489,8
4	Pública	Alegre	454,5
5	Pública	Alfredo Chaves	432,7
6	Pública	Alto Rio Novo	447,4
7	Pública	Anchieta	494,3
8	Pública	Apiacá	431,6
9	Pública	Aracruz	504,9
10	Pública	Atílio Vivácqua	450,0
11	Pública	Baixo Guandu	526,9
12	Pública	Barra de São Francisco	470,4
13	Pública	Boa Esperança	487,3
14	Pública	Bom Jesus do Norte	403,5
15	Pública	Brejo do Santo	422,2

Mediana global = 479,8

# Tabela de contingência

Tipo de Escola	Acima	Abaixo	Total
Particular	15	4	19
Pública	33	44	77
<b>Total</b>	<b>48</b>	<b>48</b>	<b>96</b>

# Estimadores de máxima verossimilhança

Estimador	Estimativa
	0.789
	0.428
	0.5



# Teste da razão de verossimilhanças

```
1 # Log-verossimilhança sob H0
2 log_L0 <- sum(tabela$acima) * log(p_hat_geral) +
3   (sum(tabela$total) - sum(tabela$acima)) * log(1 - p_hat_geral)
4
5 # Log-verossimilhança sob H1
6 log_L1 <- tabela$acima[1] * log(p_hat_particular) +
7   (tabela$total[1] - tabela$acima[1]) * log(1 - p_hat_particular) +
8   tabela$acima[2] * log(p_hat_publica) +
9   (tabela$total[2] - tabela$acima[2]) * log(1 - p_hat_publica)
10
11 Lambda <- -2 * (log_L0 - log_L1)
12 p_valor <- pchisq(Lambda, df = 1, lower.tail = FALSE)
13
14 cat("Estatística do TRV =", round(Lambda, 4), "\np-valor =", round(p_valor,
```

Estatística do TRV = 8.3596

p-valor = 0.0038

# Comparação com o teste tradicional

```
1 teste_chi2 <- chisq.test(matrix(c(tabela$acima, tabela$abaixo), nrow = 2),
2 print(teste_chi2)
```

Pearson's Chi-squared test

```
data:  matrix(c(tabela$acima, tabela$abaixo), nrow = 2)
X-squared = 7.9398, df = 1, p-value = 0.004836
```

Método	Estatística de Teste	P-valor
TRV	8.3596	0.0038
	7.9398	0.0048

