

# 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 Testes Mais Poderosos
- 2 Exemplos
- 3 Teorema de Neyman-Pearson

Sejam  $X_1, \dots, X_n$  variáveis aleatórias iid de uma distribuição dependendo de um vetor de parâmetros  $\theta \in \Omega$ . Assuma que  $\theta \in W_0$  ou  $\theta \in W_1$ , com  $W_0 \cap W_1 = \emptyset$  e  $W_0 \cup W_1 = \Omega$ . Com isso, definimos as hipóteses:

$$H_0 : \theta \in W_0 \text{ contra } H_1 : \theta \in W_1.$$

O teste de  $H_0$  contra  $H_1$  é baseado na amostra  $X_1, \dots, X_n$ , considere um subconjunto (dependendo da amostra)  $\mathcal{C}$  de  $\mathcal{S}$ , em que  $\mathcal{S}$  é o suporte da amostra aleatória. Essa região  $\mathcal{C}$  é conhecida como região crítica e sua correspondente regra de decisão é:

- Rejeite  $H_0$  (Aceite  $H_1$ ) se  $(X_1, \dots, X_n) \in \mathcal{C}$ .
- Aceita  $H_0$  (Rejeite  $H_1$ ) se  $(X_1, \dots, X_n) \notin \mathcal{C}$  ( $\in \mathcal{C}^c$ ).

O teste de  $H_0$  contra  $H_1$  é baseado na amostra  $X_1, \dots, X_n$ , considere um subconjunto (dependendo da amostra)  $\mathcal{C}$  de  $\mathcal{S}$ , em que  $\mathcal{S}$  é o suporte da amostra aleatória. Essa região  $\mathcal{C}$  é conhecida como região crítica e sua correspondente regra de decisão é:

- Rejeite  $H_0$  (Aceite  $H_1$ ) se  $(X_1, \dots, X_n) \in \mathcal{C}$ .
- Aceita  $H_0$  (Rejeite  $H_1$ ) se  $(X_1, \dots, X_n) \notin \mathcal{C}$  ( $\in \mathcal{C}^c$ ).

O erro tipo I ocorre se  $H_0$  é rejeitada quando ela é verdadeira, enquanto o erro tipo II ocorre se  $H_0$  é aceita quando  $H_1$  é verdadeira. Nível de significância é o erro do tipo I, isto é,

$$\alpha = \max_{\theta \in \mathcal{W}_0} P_{\theta}[(X_1, \dots, X_n) \in \mathcal{C}].$$

Restrito a testes de tamanho  $\alpha$ , queremos selecionar testes que minimizam o erro do tipo II, que é equivalente a maximizar a função poder. A função poder é definida por,

$$\gamma_{\mathcal{C}}(\theta) = P_{\theta}[(X_1, \dots, X_n) \in \mathcal{C}], \theta \in W_1$$

## Definição 1

Seja  $\mathcal{C}$  um subconjunto do suporte de  $(X_1, \dots, X_n)$ . Dizemos que  $\mathcal{C}$  é a melhor região crítica de tamanho  $\alpha$  se,

- a)  $\alpha = \max_{\theta \in W_0} P((X_1, \dots, X_n) \in \mathcal{C})$
- b) Para qualquer outra região  $A$  com  $\alpha = P_{\theta \in W_0}((X_1, \dots, X_n) \in A)$ ,  
 $P_{\theta}((X_1, \dots, X_n) \in \mathcal{C}) \geq P_{\theta}((X_1, \dots, X_n) \in A)$  quando  $\theta \in W_1$

Considere  $X \stackrel{\text{iid}}{\sim} \text{Binomial}(5, \theta)$ . Seja  $f(x; \theta)$  a função de probabilidade de  $X$  e considere  $H_0 : \theta = \frac{1}{2}$  e  $H_1 : \theta = \frac{3}{4}$ . Além disso, considere:



Considere  $X \stackrel{\text{iid}}{\sim} \text{Binomial}(5, \theta)$ . Seja  $f(x; \theta)$  a função de probabilidade de  $X$  e considere  $H_0 : \theta = \frac{1}{2}$  e  $H_1 : \theta = \frac{3}{4}$ . Além disso, considere:

$x$	0	1	2	3	4	5
$f(x; \frac{1}{2})$	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$
$f(x; \frac{3}{4})$	$\frac{1}{1024}$	$\frac{15}{1024}$	$\frac{90}{1024}$	$\frac{270}{1024}$	$\frac{405}{1024}$	$\frac{243}{1024}$
$\frac{f(x; \frac{1}{2})}{f(x; \frac{3}{4})}$	32	$\frac{32}{3}$	$\frac{32}{9}$	$\frac{32}{27}$	$\frac{32}{81}$	$\frac{32}{243}$

Considere o nível de significância do teste como  $\alpha = \frac{1}{32}$ . Buscamos uma melhor região crítica de tamanho  $\alpha = \frac{1}{32}$ . Se  $A_1 = \{x : x = 0\}$  ou  $A_2 = \{x : x = 5\}$ , então  $P_{\{\theta=\frac{1}{2}\}}(X \in A_1) = P_{\{\theta=\frac{1}{2}\}}(X \in A_2) = \frac{1}{32}$  e não há outro subconjunto  $A_3$  do espaço  $\{x : x = 0, 1, 2, 3, 4, 5\}$  tal que  $P_{\{\theta=\frac{1}{2}\}}(X \in A_3) = \frac{1}{32}$ . Portanto, ou  $A_1$  ou  $A_2$  é a melhor região crítica  $C$  de tamanho  $\alpha = \frac{1}{32}$  para testar  $H_0$  contra  $H_1$ .

Observamos que  $P_{\{\theta=\frac{1}{2}\}}(X \in A_1) = \frac{1}{32}$  e  $P_{\{\theta=\frac{3}{4}\}}(X \in A_1) = \frac{1}{1024}$ . Assim, se o conjunto  $A_1$  for usado como região crítica de tamanho  $\alpha = \frac{1}{32}$ , temos a situação inaceitável de que a probabilidade de rejeitar  $H_0$  quando  $H_1$  é verdadeira ( $H_0$  é falsa) é muito menor do que a probabilidade de rejeitar  $H_0$  quando  $H_0$  é verdadeira.

Por outro lado, se o conjunto  $A_2$  for usado como região crítica, então  $P_{\{\theta=\frac{1}{2}\}}(X \in A_2) = \frac{1}{32}$  e  $P_{\{\theta=\frac{3}{4}\}}(X \in A_2) = \frac{243}{1024}$ . Ou seja, a probabilidade de rejeitar  $H_0$  quando  $H_1$  é verdadeira é muito maior do que a probabilidade de rejeitar  $H_0$  quando  $H_0$  é verdadeira. Certamente, esta é uma situação mais desejável, e na verdade,  $A_2$  é a melhor região crítica de tamanho  $\alpha = \frac{1}{32}$ . Esta última afirmação decorre do fato de que quando  $H_0$  é verdadeira, existem apenas dois subconjuntos,  $A_1$  e  $A_2$ , do espaço amostral, cada um com medida de probabilidade igual a  $\frac{1}{32}$ , e do fato de que  $\frac{243}{1024} = P_{\{\theta=\frac{3}{4}\}}(X \in A_2) > P_{\{\theta=\frac{3}{4}\}}(X \in A_1) = \frac{1}{1024}$ .

Deve ser observado neste problema que a melhor região crítica  $C = A_2$  de tamanho  $\alpha = \frac{1}{32}$  é encontrada incluindo em  $C$  o ponto (ou pontos) em que  $\frac{f(x; \frac{1}{2})}{f(x; \frac{3}{4})}$  é mínimo. Assim, a razão  $\frac{f(x; \frac{1}{2})}{f(x; \frac{3}{4})}$ , que é dada na última linha da tabulação acima, nos fornece uma ferramenta precisa para encontrar uma melhor região crítica  $C$  para determinados valores dados de  $\alpha$ .

Para ilustrar o último slide, suponha  $\alpha = \frac{6}{32}$ . Quando  $H_0$  é verdadeira, cada um dos subconjuntos  $\{x : x = 0, 1\}$ ,  $\{x : x = 0, 4\}$ ,  $\{x : x = 1, 5\}$ ,  $\{x : x = 4, 5\}$  tem medida de probabilidade  $\frac{6}{32}$ . Por cálculo direto, é encontrado que a melhor região crítica desse tamanho é  $\{x : x = 4, 5\}$ . Isso reflete o fato de que a razão  $\frac{f(x; \frac{1}{2})}{f(x; \frac{3}{4})}$  tem seus dois valores mínimos em  $x = 4$  e  $x = 5$ . O poder deste teste, que tem  $\alpha = \frac{6}{32}$ , é  $P_{\{\theta = \frac{3}{4}\}}(X = 4, 5) = \frac{405}{1024} + \frac{243}{1024} = \frac{648}{1024}$ .

# Teorema de Neyman-Pearson

## Teorema 1

Sejam  $X_1, \dots, X_n$  uma amostra aleatória de uma distribuição dependendo de um vetor de parâmetros  $\theta$ . Denote,  $L(\theta, X) = \prod_{i=1}^n f(x_i, \theta)$  a função de verossimilhança e  $\underline{x} = (x_1, \dots, x_n)^\top$ . Sejam  $\theta'$  e  $\theta''$  dois valores distintos de  $\theta$ ,  $\Omega = \{\theta', \theta''\}$ , e  $k$  um número positivo, Seja  $\mathcal{C}$  tal que,

- a)  $\frac{L(\theta', \underline{x})}{L(\theta'', \underline{x})} \leq k$ , para cada  $\underline{x} \in \mathcal{C}$ ,  $\underline{X} = (X_1, \dots, X_n)^\top$
- b)  $\frac{L(\theta', \underline{x})}{L(\theta'', \underline{x})} \geq k$ , para cada  $\underline{x} \in \mathcal{C}^c$
- c)  $\alpha = P_{\theta'}(\underline{X} \in \mathcal{C})$ .

Então,  $\mathcal{C}$  é a melhor região crítica de tamanho  $\alpha$  para testar as hipóteses  $H_0 : \theta = \theta'$  contra  $H_1 : \theta = \theta''$ .

É importante entender que deseja-se mostrar que, satisfeita as condições do teorema,  $\mathcal{C}$  é a melhor região crítica de tamanho  $\alpha$ . Isso pode ser escrito matematicamente como  $P_{\theta''}(\tilde{X} \in \mathcal{C}) > P_{\theta''}(\tilde{X} \in \mathcal{A})$  para qualquer outra região crítica  $\mathcal{A}$  de tamanho  $\alpha$ .



# Demonstração

É importante entender que deseja-se mostrar que, satisfeita as condições do teorema,  $\mathcal{C}$  é a melhor região crítica de tamanho  $\alpha$ . Isso pode ser escrito matematicamente como  $P_{\theta''}(\tilde{X} \in \mathcal{C}) > P_{\theta''}(\tilde{X} \in \mathcal{A})$  para qualquer outra região crítica  $\mathcal{A}$  de tamanho  $\alpha$ .

Ou seja, queremos mostrar que

$$P_{\theta''}(\tilde{X} \in \mathcal{C}) - P_{\theta''}(\tilde{X} \in \mathcal{A}) > 0.$$

# Demonstração

Consideramos o caso em que  $X_1, \dots, X_n$  são variáveis contínuas (o caso discreto é análogo).

**Notação Simplificada:**  $\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} L(\theta, \underline{X}) d\underline{X} \equiv \int L(\theta)$

# Demonstração

Consideramos o caso em que  $X_1, \dots, X_n$  são variáveis contínuas (o caso discreto é análogo).

**Notação Simplificada:**  $\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} L(\theta, \underline{X}) d\underline{X} \equiv \int L(\theta)$

Queremos mostrar que,

$$\int_{\mathcal{C}} L(\theta'') - \int_A L(\theta'') > 0, \text{ em que } A \text{ tem tamanho } \alpha.$$

Note que  $\mathcal{C} = (\mathcal{C} \cap A) \cup (\mathcal{C} \cap A^c)$  e  $A = (\mathcal{C} \cap A) \cup (\mathcal{C}^c \cap A)$ .

Logo,

$$\begin{aligned}\int_C L(\theta'') - \int_A L(\theta'') &= \int_{C \cap A} L(\theta'') + \int_{C \cap A^c} L(\theta'') - \int_{C \cap A} L(\theta'') - \int_{C^c \cap A} L(\theta'') \\ &= \int_{C \cap A^c} L(\theta'') - \int_{C^c \cap A} L(\theta'')\end{aligned}$$

Segue que,

$$\int_{C \cap A^c} L(\theta'') - \int_{C^c \cap A} L(\theta'') \geq \int_{C \cap A^c} \frac{1}{k} L(\theta') - \int_{C^c \cap A} \frac{1}{k} L(\theta')$$

Usando a parte a e b do teorema

$$\begin{aligned} &= \frac{1}{k} \left( \int_{C \cap A^c} L(\theta') - \int_{C^c \cap A} L(\theta') \right) \\ &= \frac{1}{k} \left( \int_{C \cap A^c} L(\theta') + \int_{C \cap A} L(\theta') - \right. \\ &\quad \left. - \int_{C \cap A} L(\theta') - \int_{C^c \cap A} L(\theta') \right) \\ &= \frac{1}{k} \left( \int_C L(\theta') - \int_A L(\theta') \right) \Rightarrow \text{Sob } H_0 \\ &= \frac{1}{k} (\alpha - \alpha) = 0 \end{aligned}$$

# Exemplo

$$X_1, \dots, X_n \stackrel{\text{iid}}{\sim} N(\theta, 1)$$

$$H_0 : \theta = \theta' = 0$$

$$H_1 : \theta = \theta'' = 1, \quad L(\theta, \underline{X}) = \left( \frac{1}{\sqrt{2\pi}} \right)^n \exp \left\{ -\frac{1}{2} \sum (X_i - \theta)^2 \right\}$$

$$\frac{L(\theta', \underline{x})}{L(\theta'', \underline{x})} \leq k,$$

$$\begin{aligned} \frac{L(\theta', \underline{x})}{L(\theta'', \underline{x})} &= \exp \left\{ -\sum_{i=1}^n x_i + \frac{n}{2} \right\} \leq k \\ \Rightarrow -\sum x_i + \frac{n}{2} &\leq k' = \log k \\ \Rightarrow \underbrace{\sum x_i}_{\text{Região Crítica}} &\geq k'' \end{aligned}$$

Portanto, a melhor região crítica  $C$  é o conjunto de todas as amostras  $(x_1, x_2, \dots, x_n)$  para as quais a soma dos  $x_i$  é maior ou igual a  $k''$ . O valor de  $k''$  pode ser determinado de modo que o tamanho da região crítica seja igual ao nível de significância desejado  $\alpha$ . Essa região crítica é dada por:

$$C = \{(x_1, x_2, \dots, x_n) : \sum_{i=1}^n x_i \geq k''\}.$$

O teste pode ser baseado na estatística  $\bar{X}$ , pois  $\sum_{i=1}^n x_i \geq k''$  é equivalente a  $\bar{X} \geq c_1$ . Se  $H_0$  for verdadeira, ou seja,  $\theta = \theta_0 = 0$ , então  $\bar{X}$  segue uma distribuição  $N(0, 1/n)$ . Dado um nível de significância  $\alpha$ , podemos calcular  $c_1$  em R como  $c_1 = qnorm(1 - \alpha, 0, 1/\sqrt{n})$ .



Portanto, se os valores experimentais de  $X_1, X_2, \dots, X_n$  forem, respectivamente,  $x_1, x_2, \dots, x_n$ , podemos calcular  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ . Se  $\bar{x} \geq c_1$ , a hipótese simples  $H_0 : \theta = \theta_0 = 0$  será rejeitada no nível de significância  $\alpha$ ; caso contrário, a hipótese  $H_0$  será aceita. A probabilidade de rejeitar  $H_0$  quando  $H_0$  é verdadeira é igual a  $\alpha$ , o nível de significância. A probabilidade de rejeitar  $H_0$  quando  $H_0$  é falsa, ou seja, o valor de poder do teste quando  $\theta = \theta_1 = 1$ , pode ser calculada como indicado na equação fornecida.

$$P_{H1}(\bar{X} \geq c_1) = \int_{c_1}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1/n}} \exp\left(-\frac{(\bar{x} - 1)^2}{2(1/n)}\right) d\bar{x}.$$

Portanto, se os valores experimentais de  $X_1, X_2, \dots, X_n$  forem, respectivamente,  $x_1, x_2, \dots, x_n$ , podemos calcular  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ . Se  $\bar{x} \geq c_1$ , a hipótese simples  $H_0 : \theta = \theta_0 = 0$  será rejeitada no nível de significância  $\alpha$ ; caso contrário, a hipótese  $H_0$  será aceita. A probabilidade de rejeitar  $H_0$  quando  $H_0$  é verdadeira é igual a  $\alpha$ , o nível de significância. A probabilidade de rejeitar  $H_0$  quando  $H_0$  é falsa, ou seja, o valor de poder do teste quando  $\theta = \theta_1 = 1$ , pode ser calculada como indicado na equação fornecida.

$$P_{H1}(\bar{X} \geq c_1) = \int_{c_1}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1/n}} \exp\left(-\frac{(\bar{x} - 1)^2}{2(1/n)}\right) d\bar{x}.$$

Por exemplo, se  $n = 25$  e  $\alpha$  for 0.05, então  $c_1 = \text{qnorm}(0.95, 0, 1/5) = 0.329$ , usando R. Portanto, o poder do teste para detectar  $\theta = 1$ , é calculado por  $1 - \text{pnorm}(0.329, 1, 1/5) = 0.9996$ .

## Exercício 8.1.2

Vamos considerar o problema de testar a hipótese simples  $H_0 : \theta = \theta_0 = 2$  contra a hipótese alternativa simples  $H_1 : \theta = \theta_1 = 4$  para a variável aleatória  $X$ , que possui a função de densidade de probabilidade (pdf) dada por  $f(x; \theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$ , onde  $0 < x < \infty$  e zero caso contrário. Seja  $X_1$  e  $X_2$  representam uma amostra aleatória de tamanho 2 desta distribuição. Mostre que o melhor teste de  $H_0$  contra  $H_1$  é usando a estatística  $X_1 + X_2$ !

A razão de verossimilhança é dada por:

$$\begin{aligned}\frac{L(\theta_0; X_1, X_2)}{L(\theta_1; X_1, X_2)} &= \frac{\frac{1}{\theta_0} e^{-\frac{X_1}{\theta_0}} \frac{1}{\theta_0} e^{-\frac{X_2}{\theta_0}}}{\frac{1}{\theta_1} e^{-\frac{X_1}{\theta_1}} \frac{1}{\theta_1} e^{-\frac{X_2}{\theta_1}}} \\ &= 4e^{-\left(\frac{X_1 + X_2}{4}\right)},\end{aligned}$$

em que utilizamos as hipóteses  $H_0 : \theta = 2$  e  $H_1 : \theta = 4$ . Agora, queremos encontrar a melhor região crítica  $C$  de tamanho  $\alpha$  para testar  $H_0$  contra  $H_1$ . Neste caso,  $\alpha$  representa o nível de significância do teste.

Usando a razão de verossimilhança, temos que:

$$\frac{L(\theta_0; X_1, X_2)}{L(\theta_1; X_1, X_2)} \leq k \quad \text{se, e somente se,} \quad 4e^{-\left(\frac{X_1 + X_2}{4}\right)} \leq k,$$

ou seja,  $X_1 + X_2 \geq k_3$ . Portanto, a região crítica  $C$  pode ser definida como:

$$C = \{(X_1, X_2) : X_1 + X_2 \geq k_3\},$$

em que  $k$  é escolhido de forma a controlar o nível de significância  $\alpha$ . O teste de hipóteses consiste em verificar se a amostra cai dentro ou fora da região crítica  $C$ . Portanto, o melhor teste usa a estatística  $X_1 + X_2$  e a melhor região crítica  $C$  é definida por  $C = \{(x_1, x_2) / x_1 + x_2 \geq k\}$ .

# Para

- Exercícios da seção 8.1: 2, 3, 5, 6, 8, 9, 10.

# Referências I



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