# 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

# Propriedades de Uma Estatística Suficiente

Seja que $X_1, X_2, \ldots, X_n$ seja uma amostra aleatória de uma variável aleatória com função de densidade ou função de massa de probabilidade $f(x;\theta)$, $\theta \in \Omega$. **Observemos que uma estatística suficiente não é única**. Seja $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ uma estatística suficiente para $\theta$. Denote por $f_{Y_1}(y,\theta)$ a fdp de $Y_1$. Seja $Y_2 = g(Y_1)$ uma estatística, em que $g$ é uma função bijetora, então pela definição de Estatística Suficiente temos,

$$\frac{f(x_1;\theta)f(x_2;\theta)\ldots f(x_n;\theta)}{f_{Y_1}(y,\theta)} = H(x_1, x_2, \ldots, x_n)$$

$$\Rightarrow f(x_1;\theta)f(x_2;\theta)\ldots f(x_n;\theta) = f_{Y_1}(y,\theta)H(x_1, x_2, \ldots, x_n)$$

Pelo teorema de Neyman (Critério da fatoração),

$$\begin{aligned} f(x_1;\theta)f(x_2;\theta)\dots f(x_n;\theta) &= k_1(y_1,\theta)k_2(x_1,x_2,\dots,x_n) \\ &= k_1(g^{-1}(y_2),\theta)k_2(x_1,x_2,\dots,x_n) \\ &= k_1^*(y_2,\theta)k_2(x_1,x_2,\dots,x_n), \end{aligned}$$

logo $Y_2$ também é uma estatística suficiente para $\theta$.

# Propriedades Básicas

Das propriedades básicas de Esperança Condicional, segue que, se $X_1$ e $X_2$ são variáveis aleatórias tais que a variância de $X_2$ existe, então

$$\begin{aligned} E[X_2] &= E[E(X_2|X_1)] \\ \text{Var}(X_2) &= \text{Var}[E(X_2|X_1)] + \underbrace{E[\text{Var}(X_2|X_1)]}_{\geq 0} \\ \Rightarrow \text{Var}(X_2) &\geq \text{Var}[E(X_2|X_1)]. \end{aligned}$$

# Propriedades Básicas

vamos considerar a estatística suficiente $Y_1$ como $X_1$ e $Y_2$, uma estatística não-viesada de $\theta$, como $X_2$. Assim, $X_2$ é um estimador não viesado de $\theta$ e $X_1$ uma estatística suficiente para $\theta$. Defina, $\xi(y_1) = E(Y_2|Y_1 = y_1)$. Logo, podemos escrever $\xi(Y_1) = E(Y_2|Y_1)$ e,

$$E[\xi(Y_1)] = E(E(Y_2|Y_1)) = E(Y_2) = \theta$$
$$\text{Var}[\xi(Y_1)] = \text{Var}(E(Y_2|Y_1)) \leq \text{Var}(Y_2).$$

Esses resultados podem ser enunciados como um teorema.

# Teorema de Rao-Blackwell

## Teorema 1

*Seja* $X_1, X_2, \ldots, X_n$, *uma amostra aleatória de uma distribuição (contínua ou discreta) com função de densidade ou função de massa de probabilidade* $f(x;\theta)$, $\theta \in \Omega$. *Seja* $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ *uma estatística suficiente para* $\theta$ *e* $Y_2 = u_2(X_1, X_2, \ldots, X_n)$ *um estimador não-viesado de* $\theta$. *Considere, ainda,* $E(Y_2|y_1) = \xi(y_1)$. *Então* $\xi(Y_1)$ *é um estimador não-viesado de* $\theta$ *e sua variância é menor ou igual à de* $Y_2$.

# Teorema de Rao-Blackwell

## Teorema 1

*Seja* $X_1, X_2, \ldots, X_n$, *uma amostra aleatória de uma distribuição (contínua ou discreta) com função de densidade ou função de massa de probabilidade* $f(x;\theta)$, $\theta \in \Omega$. *Seja* $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ *uma estatística suficiente para* $\theta$ *e* $Y_2 = u_2(X_1, X_2, \ldots, X_n)$ *um estimador não-viesado de* $\theta$. *Considere, ainda,* $E(Y_2|y_1) = \xi(y_1)$. *Então* $\xi(Y_1)$ *é um estimador não-viesado de* $\theta$ *e sua variância é menor ou igual à de* $Y_2$.

## Demonstração

Ver slide anterior de propriedades básicas!

Este teorema nos diz que, na busca por um EMVU (Estimador de Mínima Variância Não-viesado) de um parâmetro, podemos restringir essa busca a funções da estatística suficiente. Se começarmos com um estimador não-viesado $Y_2$, podemos sempre melhorá-lo calculando $E(Y_2|y_1) = \xi(y_1)$, de modo que $\xi(Y_1)$ seja um estimador não-viesado com uma variância menor que a de $Y_2$.

# Teorema

## Teorema 2

*Sejam $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma distribuição com função de densidade ou função de massa de probabilidade $f(x;\theta), \theta \in \Omega$. Se uma estatística suficiente $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ para $\theta$ existe e se um estimador de máxima verossimilhança $\hat{\theta}$ para $\theta$ também existe de forma única, então $\hat{\theta}$ é uma função da estatística suficiente $Y_1 = u_1(X_1, X_2, \ldots, X_n)$.*

# Demonstração

Seja $f_{Y_1}(y_1; \theta)$ a função de densidade ou função de massa de probabilidade de $Y_1$. Então, pela definição de suficiência (teorema de Neyman), a função de verossimilhança

$$\begin{aligned} L(\theta) &= f(x_1; \theta) f(x_2; \theta) \dots f(x_n; \theta) \\ &= f_{Y_1}[u_1(x_1, x_2, \dots, x_n); \theta] H(x_1, x_2, \dots, x_n) \\ &= k_1(y_1, \theta) k_2(x_1, x_2, \dots, x_n) \end{aligned}$$

Assim, $L$ e $k_1(y_1, \theta)$, como funções de $\theta$, são maximizadas simultaneamente. Por hipótese, o EMV $\hat{\theta}$ existe e é único, logo, há apenas um valor de $\theta$ que maximiza $L$ e, portanto, $k_1[u_1(x_1, x_2, \dots, x_n); \theta]$, esse valor de $\theta$ deve ser uma função de $u_1(x_1, x_2, \dots, x_n)$. Portanto, o EMV $\hat{\theta}$ é uma função da estatística suficiente $Y_1 = u_1(X_1, X_2, \dots, X_n)$.

Sabemos, de aulas anteriores, que, em geral, os EMVs são estimadores assintoticamente não-viesados de $\theta$. Portanto, uma maneira de proceder é encontrar uma estatística suficiente e, em seguida, encontrar o EMV. Com base nisso, muitas vezes podemos obter um estimador não-viesado que é uma função da estatística suficiente. Esse processo é ilustrado no exemplo a seguir.

# Exemplo

Sejam $X_1, \ldots, X_n \overset{\text{iid}}{\sim} exp(\theta)$.
Suponha que desejamos um EMVU (Estimador de Variância Mínima Não-viesado) para $\theta$.

$$\ell(\theta) = n \log \theta - \theta \sum_{i=1}^{n} x_i \tag{1}$$

$\hat{\theta} = \dfrac{1}{\bar{X}}$ é o EMV de $\theta$. Também podemos verificar que $\bar{X}$ é uma estatística suficiente para $\theta$.

Note que $\hat{\theta} = \frac{n}{Y_1}$ é uma função da estatística suficiente $Y_1$. Além disso, como $\hat{\theta}$ é o EMV de $\theta$, ele é assintoticamente não-viesado. Portanto, como um primeiro passo, vamos determinar sua esperança. Neste problema, $X_i$ são variáveis aleatórias independentes e identicamente distribuídas com distribuição Gama $\Gamma(1, 1/\theta)$; portanto, $Y_1 = \sum_{i=1}^{n} X_i$ segue uma distribuição $\Gamma(n, 1/\theta)$. Assim,

$$E(\hat{\theta}) = E\left(\frac{n}{Y_1}\right) = nE\left(\frac{1}{\sum_{i=1}^{n} X_i}\right)$$

$$
\begin{aligned}
nE\left(\frac{1}{\sum_{i=1}^{n} X_i}\right) &= n\int_0^{\infty} \frac{\left(\frac{1}{\theta}\right)^n}{\Gamma(n)} t^{-1} t^{n-1} e^{-\frac{t}{\theta}}\, dt \\
&= n\int_0^{\infty} \frac{\left(\frac{1}{\theta}\right)^n}{\Gamma(n)} t^{(n-1)-1} e^{-\frac{t}{\theta}}\, dt \\
&= \frac{n}{\theta^n \Gamma(n)} \int_0^{\infty} \underbrace{t^{(n-1)-1} e^{-\frac{t}{\theta}}}_{\text{Núcleo da densidade de uma } \Gamma(n-1,\theta)} dt \\
&= \frac{n}{\theta^n \Gamma(n)} \int_0^{\infty} \frac{\left(\frac{1}{\theta}\right)^{(n-1)}}{\Gamma(n-1)} t^{(n-1)-1} e^{-\frac{t}{\theta}} \frac{\Gamma(n-1)}{\left(\frac{1}{\theta}\right)^{(n-1)}} dt \\
&= \frac{n\Gamma(n-1)}{\frac{\theta^n}{\theta^{(n-1)}}\Gamma(n)} = \theta \frac{n}{n-1}
\end{aligned}
$$

Portanto, o estimador $\tilde{\theta} = \hat{\theta}\frac{(n-1)}{n}$ é um estimador não viesado de variância mínima para $\theta$.

**Observação:** Como o estimador não-viesado $\xi(Y_1)$, em que $\xi(Y_1) = E(\hat{\theta}|y_1)$, possui uma variância menor do que o estimador não-viesado $Y_2$ de $\theta$, às vezes, raciocinamos da seguinte maneira. Seja a função $\Upsilon(y_3) = E[\xi(Y_1)|Y_3 = y_3]$, onde $Y_3$ é uma estatística que não é suficiente para $\theta$. Pelo teorema de Rao-Blackwell, temos $E[\Upsilon(Y_3)] = \theta$ e $\Upsilon(Y_3)$ possui uma variância menor do que $\xi(Y_1)$. Consequentemente, $\Upsilon(Y_3)$ deve ser melhor do que $\xi(Y_1)$ como um estimador não-viesado de $\theta$. No entanto, isso não é verdade, porque $Y_3$ não é suficiente; assim, $\theta$ está presente na distribuição condicional de $Y_1$, dado $Y_3 = y_3$, e na média condicional $\Upsilon(y_3)$. Embora de fato $E[\Upsilon(Y_3)] = \theta$, $\Upsilon(Y_3)$ nem sequer é uma estatística, pois envolve o parâmetro desconhecido $\theta$ e, portanto, não pode ser usado como uma estimativa.

# Completude

## Definição 1

*Seja a variável aleatória* $Z$ *do tipo contínuo ou discreto, com função de densidade ou massa de probabilidade que pertence à família* $\{h(z;\theta), \theta \in \Omega\}$. *Se a condição* $E[u(Z)] = 0$, *para todo* $\theta \in \Omega$, *implica que* $u(z)$ *seja igual a zero, exceto em um conjunto de pontos com probabilidade zero, então para cada* $h(z;\theta)$, $\theta \in \Omega$, *a família* $\{h(z;\theta) : \theta \in \Omega\}$ *é chamada de família completa de funções de densidade ou massa de probabilidade.*

# Exemplo 1

$X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Poisson}(\theta)$. Sabemos que $\sum_{i=1}^{n} X_i$ é uma estatística suficiente para $\theta$ e $Z = \sum_{i=1}^{n} X_i \sim \text{Poisson}(n\theta)$, $\Omega = (0, +\infty)$. Seja $u(\cdot)$ tal que $E(u(Z)) = 0, \; \forall \theta > 0$.

$$0 = E(u(Z)) = \sum_{z=0}^{+\infty} u(z) \frac{(n\theta)^z}{z!} e^{-n\theta}$$

$$\Rightarrow 0 = \sum_{z=0}^{+\infty} u(z) \frac{n^z}{z!} \theta^z, \forall \theta > 0$$

$$\Rightarrow \frac{u(z) n^z}{z!} = 0, \; \forall z \in \{0, 1, \ldots\} \Rightarrow u(z) = 0, \forall z \in \{0, 1, \ldots\}$$

# Exemplo 2

Considere a família $\{h(z,\theta); \theta > 0\}$ dada por

$$h(z;\theta) = \frac{1}{\theta} e^{-\left(\frac{z}{\theta}\right)}, \;\theta > 0, \text{ ou seja, } Z \sim \exp(\theta).$$

# Exemplo 2

Considere a família $\{h(z,\theta); \theta > 0\}$ dada por

$$h(z;\theta) = \frac{1}{\theta} e^{-\left(\frac{z}{\theta}\right)}, \; \theta > 0, \text{ ou seja, } Z \sim \exp(\theta).$$

$$E(u(Z)) = 0, \forall \theta > 0 \Rightarrow \int_0^{+\infty} u(z) \frac{1}{\theta} e^{-\left(\frac{z}{\theta}\right)} dz = 0, \forall \theta > 0$$
$$\Rightarrow \int_0^{+\infty} u(z) e^{-\left(\frac{z}{\theta}\right)} dz = 0, \forall \theta > 0$$

Notem que $\ell(\theta) = \int_0^{+\infty} u(z) e^{-\left(\frac{z}{\theta}\right)} dz = 0$ é a transformada de Laplace de $u(\cdot)$. Assim, $\ell(\theta) = 0 \Rightarrow u(z) = 0, \forall z > 0$.

Seja $Y1 = u_1(X_1, X_2, ..., X_n)$ uma estatística suficiente para $\theta$. Tome $\xi(Y_1)$ tal que $E(\xi(Y_1)) = \theta$. Seja $\psi(Y_1)$ outra função de $Y_1$ de forma que, também tenhamos $E[\psi(Y_1)] = \theta$ para todos os valores de $\theta$, $\theta \in \Omega$. Portanto, $E[\xi(Y_1) - \psi(Y_1)] = 0$, $\theta \in \Omega$. Se a família de $Y_1$ é completa, $\xi(Y_1) - \psi(Y_1) = 0$, ou seja, $\xi(Y_1) = \psi(Y_1)$. Assim, $\xi(Y_1)$ é a única (com prob=1) função de $Y_1$ que é um estimador não viesado para $\theta$.

Seja $Y1 = u_1(X_1, X_2, ..., X_n)$ uma estatística suficiente para $\theta$. Tome $\xi(Y_1)$ tal que $E(\xi(Y_1)) = \theta$. Seja $\psi(Y_1)$ outra função de $Y_1$ de forma que, também tenhamos $E[\psi(Y_1)] = \theta$ para todos os valores de $\theta$, $\theta \in \Omega$. Portanto, $E[\xi(Y_1) - \psi(Y_1)] = 0$, $\theta \in \Omega$. Se a família de $Y_1$ é completa, $\xi(Y_1) - \psi(Y_1) = 0$, ou seja, $\xi(Y_1) = \psi(Y_1)$. Assim, $\xi(Y_1)$ é a única (com prob=1) função de $Y_1$ que é um estimador não viesado para $\theta$.

A conclusão anterior juntamente com o Teorema de Rao-Blackwel garantem que $\xi(Y_1)$ é o ENVVUM de $\theta$, conforme próximo teorema.

## Teorema 3

*Seja $Y_1$ uma estatística suficiente para $\theta$ e assuma que a família de $Y_1$ é completa. Se existe uma função $\xi(\cdot)$ tal que $E(\xi(Y_1)) = \theta$, $\forall \theta \in \Omega$, então, $\xi(Y_1)$ é o ENVVUM de $\theta$.*

## Para 🏠

- **Exercícios da seção 7.3:** 1,3,4,5,6
- **Exercícios da seção 7.4:** 1,2,3,4,5,7,8,9

# Referências I


- Bolfarine, Heleno e Mônica Carneiro Sandoval (2001). *Introdução à inferência estatística*. Vol. 2. SBM.
- Casella, George e Roger L Berger (2021). *Statistical inference*. Cengage Learning.
- Hogg, RV, J McKean e AT Craig (2019). *Introduction to Mathematical Statistics*.