

FGV EMAp

João Pedro Jerônimo

## Inferência Estatística

Revisão para A2

Rio de Janeiro

2025

# Conteúdo

1	Distribuição Amostral de Estimadores .....	3
2	Distribuição Chi-Quadrado .....	6
2.1	Propriedades .....	7
3	Distribuição Conjunta da Média e Variância Amostral .....	9
3.1	Independência da Média e Variância Amostrais .....	10
4	Distribuições $t$ .....	13
4.1	Propriedades .....	15
5	Intervalos de Confiança .....	18
5.1	Intervalo de Confiança para a média de uma Normal .....	19
5.2	Intervalos de Confiança Unilaterais .....	20
5.3	Intervalo de confiança para outros parâmetros .....	21
6	Análise Bayesiana de Amostras Normais .....	23
6.1	Família de Conjugados .....	24
6.2	Marginais .....	24
6.3	Distribuições Impróprias .....	25
7	Estimadores não-viezados .....	26
7.1	Estimador não-viezado da Variância .....	27
7.2	Limitações .....	28
8	Análise e Teste de Hipóteses .....	29
8.1	Hipóteses Nula e Alternativa .....	30
8.2	Região Crítica e Testes Estatísticos .....	31
8.3	Função de Poder e Tipos de Erro .....	31
8.4	Induzindo um nível de significância .....	32
8.5	p-valor .....	33
8.6	Calculando p-valores .....	33
8.7	Equivalência de testes e conjuntos de confiança .....	33
9	Testes $t$ .....	35
9.1	Testando Hipóteses sobre a Média de uma Normal quando a Variância é Desconhecida .....	36
9.2	Propriedades dos testes $t$ .....	36
9.3	Teste $t$ pareado .....	38
9.4	Testando uma alternativa bilateral .....	39
10	Comparando as médias de duas Distribuições Normais .....	41
10.1	O $t$ -teste biamostral .....	42
10.2	Poder do Teste .....	43
10.3	Alternativas Bilaterais .....	44

# Distribuição Amostral de Estimadores

Quando temos um estimador  $\delta$ , e note que estamos falando de um **estimador** e não de uma **estimativa**, temos que, como ele é função de variáveis aleatórias, ele próprio é uma variável aleatória, que possui sua própria distribuição, seus próprios parâmetros, média, variância, etc. Essa distribuição própria da estimativa é o que chamamos de **Distribuição Amostral do Estimador**

**Definição 1.1** (Distribuição Amostral do Estimador): Dadas as variáveis aleatórias  $\underline{X} = (X_1, \dots, X_n)$  e  $T = r(\underline{X})$  um estimador, onde  $\underline{X}$  tem uma distribuição indexada pelo parâmetro  $\theta$ , então a distribuição de  $T|\theta$  chamada de Distribuição Amostral de  $T$ . ( $\mathbb{E}_\theta[T]$  é a média de  $T$  na distribuição amostral)

O nome vem do fato que  $T$  depende da amostra  $\underline{X}$ . Na maioria das vezes,  $T$  não depende de  $\theta$ . Mas por que essa distribuição me é interessante?

Vamos supor que eu tenho um estimador  $\hat{\theta}$  de  $\theta$ , pode me ocorrer de eu querer saber a chance de o meu estimador estar próximo do meu  $\theta$  de verdade, por exemplo, qual a chance de a distância entre meu estimador e meu  $\theta$  ser de só 0.1 medidas? Então podemos querer calcular:

$$\mathbb{P}(|\hat{\theta} - \theta| < 0.1) \quad (1)$$

Pela lei da probabilidade total, temos também:

$$\mathbb{P}(|\hat{\theta} - \theta| < 0.1) = \mathbb{E}[\mathbb{P}(|\hat{\theta} - \theta| < 0.1 | \theta)] \quad (2)$$

Outro uso que podemos derivar para a distribuição amostral é escolher entre vários experimentos qual será performedo para obter o melhor estimador de  $\theta$ . Por exemplo, podemos querer saber qual a quantidade de amostras necessárias para atingir um objetivo em específico

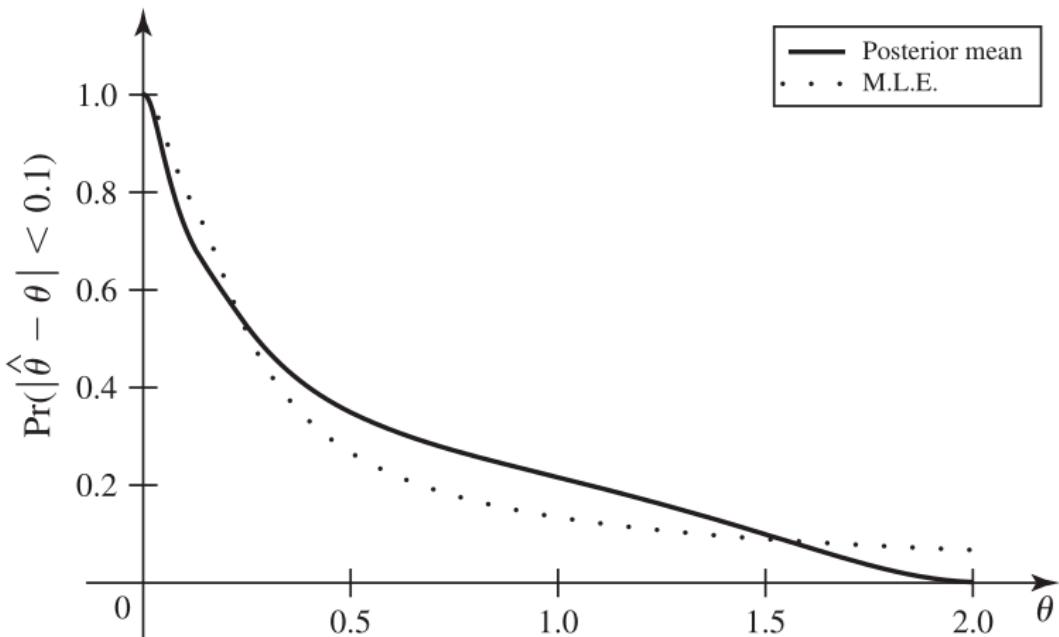


Figura 1: Imagem que representa  $\mathbb{P}(|\hat{\theta} - \theta| < 0.1)$  em função da quantidade de amostras em um dos exemplos do livro, mostrando que dependendo da nossa situação, podemos querer escolher estimadores diferentes

Uma outra medida interessante que foi apresentada em um dos exemplos do livro é uma distância relativa:

$$\mathbb{P}\left(\left|\frac{\hat{\theta}}{\theta} - 1\right| < 0.1\right) \quad (3)$$

Ou seja, a probabilidade de que meu estimador esteja a pelo menos 10% de  $\theta$  de distância de  $\theta$

*Exemplo 1.1:* Vamos tentar condensar tudo o que vimos em um exemplo. Vamos supor que temos uma clínica que está a fim de identificar ou prever pacientes candidatos a um remédio específico para tratamento da depressão. Então podemos modelar a variável aleatória de um paciente usar ou não esse remédio como uma Bernoulli com  $\theta$  de chance de utilizar o remédio ( $\mathbb{P}(X = 1) = \theta$ ). Sabemos por capítulos anteriores que  $T = \frac{1}{n} \sum_{i=1}^n X_i$  (A proporção de pacientes que vão utilizar o remédio) é uma estatística suficiente e também é o EVM (Estimador de Máxima Verossimilhança) de  $\theta$

Porém,  $T$  também é uma variável aleatória com distribuição própria, então ela pode assumir vários valores, mas queremos que ela seja o mais próximo possível de  $\theta$ . Então, que tal calcularmos:

$$\mathbb{P}(|T - \theta| < 0.1) \quad (4)$$

Para isso temos que saber exatamente a distribuição de  $T$ . Não é muito difícil, na verdade! Sabemos que  $T = \frac{1}{n}Y$  com  $Y = \sum_{i=1}^n X_i$  e  $Y|\theta \sim \text{Bin}(n, \theta)$ . Então sabemos que:

$$\begin{aligned} \mathbb{P}(T = t|\theta) &= \mathbb{P}\left(\frac{1}{n}Y = t|\theta\right) = \mathbb{P}(Y = nt|\theta) \\ &= \binom{n}{nt} \theta^{nt} (1-\theta)^{n-nt} \end{aligned} \quad (5)$$

Assim, encontramos a nossa Distribuição Amostral do estimador  $T$ . Então agora poderíamos calcular a equação mencionada anteriormente

$$\mathbb{P}(|T - \theta| < 0.1) = \mathbb{P}(-0.1 < T - \theta < 0.1) = \mathbb{P}(\theta - 0.1 < T < \theta + 0.1) \quad (6)$$

Podemos utilizar a distribuição amostral de  $T$  que encontramos anteriormente, porém, por questões de praticidade, vou fazer um pouco diferente:

$$\mathbb{P}(\theta - 0.1 < T < \theta + 0.1) = \mathbb{P}(\lceil(\theta - 0.1)n\rceil \leq Y \leq \lfloor(\theta + 0.1)n\rfloor) \quad (7)$$

Eu adicionei o piso e o teto por conta que  $Y$  assume apenas valores inteiros. Então temos que:

$$\mathbb{P}(|T - \theta| < 0.1) = \sum_{k=\lceil(\theta - 0.1)n\rceil}^{\lfloor(\theta + 0.1)n\rfloor} \binom{n}{k} \theta^k (1-\theta)^{n-k} \quad (8)$$

# **Distribuição Chi-Quadrado**

Essa distribuição é muito utilizada dentro da estatística serve como base para uma outra distribuição que veremos posteriormente. Ela vai ser útil no próximo capítulo também pois ela é a distribuição do estimador de máxima verossimilhança da variância de uma Normal com média  $\mu$  conhecida

$$\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \quad (9)$$

**Definição 2.1:**  $\forall m \in \mathbb{R}$ , uma distribuição  $\text{Gamma}\left(\frac{m}{2}, \frac{1}{2}\right)$  é também chamada de  $\chi_m^2$  (Chi quadrado com  $m$  graus de liberdade). Ou seja, se  $X \sim \chi_m^2$ :

$$f_X(x) = \frac{\left(\frac{1}{2}\right)^{\frac{m}{2}}}{\Gamma\left(\frac{m}{2}\right)} x^{\frac{m}{2}-1} e^{-\frac{1}{2}x} \quad (10)$$

## 2.1 Propriedades

**Teorema 2.1.1:** Se  $X \sim \chi_m^2$  então:

$$\mathbb{E}[X] = m \quad \mathbb{V}[X] = 2m \quad (11)$$

*Demonstração:* A esperança de uma  $\text{Gamma}(\alpha, \beta)$  é  $\frac{\alpha}{\beta}$ , logo:

$$\mathbb{E}[X] = \frac{\frac{m}{2}}{\frac{1}{2}} = m \quad (12)$$

E a variância é  $\frac{\alpha}{\beta^2}$ , logo:

$$\mathbb{V}[X] = \frac{\frac{m}{2}}{\frac{1}{4}} = 2m \quad (13)$$

□

**Teorema 2.1.2:** A função geradora de momentos de uma  $\chi_m^2$  é dada por

$$\psi(t) = \left( \frac{1}{1-2t} \right)^{m/2} \quad \left( t < \frac{1}{2} \right) \quad (14)$$

**Teorema 2.1.3:** Se  $X_1, \dots, X_k$  são iid e  $X_i \sim \chi_{m_i}^2$ , então  $Y = \sum_{j=1}^k X_j \sim \chi_{\sum_{j=1}^k m_j}^2$

*Demonstração:* Sabemos que, dado a FGM de  $X$  ( $\psi_X$ ) e de  $Y$  ( $\psi_Y$ ) onde  $X$  e  $Y$  são iid, então a FGM de  $X + Y$  é  $\psi_X \psi_Y$ . Sabendo disso, calculamos a FGM de  $X_1 + \dots + X_k$ :

$$\begin{aligned}\psi_Y(t) &= \prod_{j=1}^k \left( \frac{1}{1-2t} \right)^{m_j/2} \quad \left( t < \frac{1}{2} \right) \\ &= \left( \frac{1}{1-2t} \right)^{\frac{1}{2} \sum_{j=1}^k m_j} \quad \left( t < \frac{1}{2} \right)\end{aligned}\tag{15}$$

□

**Teorema 2.1.4:** Se  $X \sim N(0, 1)$ , então  $Y = X^2 \sim \chi_1^2$

*Demonstração:* Sabemos que se  $X$  tem pdf  $f_X(x)$  e  $Z = h(X)$ , então a pdf de  $Z$  é

$$f_Z(z) = \frac{f_X(x)}{|h'(x)|} \tag{16}$$

Então, considerando  $h(x) = x^2$  e  $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ , temos que

$$f_Z(z) = \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}}{2x} = \frac{1}{2\sqrt{\pi z}} e^{-\frac{1}{2}z} \tag{17}$$

Perceba que isso é a densidade de uma  $\chi_1^2$ , veja:

$$\frac{\left(\frac{1}{2}\right)^{\frac{1}{2}}}{\Gamma\left(\frac{1}{2}\right)} z^{\frac{1}{2}-1} e^{-\frac{1}{2}z} = \frac{\frac{1}{\sqrt{2}}}{\sqrt{\pi}} \frac{1}{\sqrt{z}} e^{-\frac{1}{2}z} = \frac{1}{\sqrt{2\pi z}} e^{-\frac{1}{2}z} \tag{18}$$

Logo,  $Z \sim \chi_1^2$ . A função  $f_Z(z)$  possui um termo  $\frac{1}{2}$  pois precisamos lembrar que a normal vai de  $[-\infty, \infty]$ , então para que  $h(x)$  seja monótona, precisamos restringir em  $[-\infty, 0]$  e  $[0, \infty]$  então obtemos duas funções que integram  $\frac{1}{2}$  em cada intervalo, logo o resultado integra 1 □

**Corolário 2.1.4.1:** Se  $X_1, \dots, X_n \sim N(0, 1)$ , então:

$$X_1^2 + \dots + X_n^2 \sim \chi_m^2 \tag{19}$$

# **Distribuição Conjunta da Média e Variância Amostral**

Vamos supor que estamos fazendo um experimento e coletamos amostras de uma distribuição normal com média  $\mu$  e variância  $\sigma^2$ , porém, não sabemos nenhum dos dois, então queremos estimá-los! Podemos escolher vários estimadores, por exemplo, o EVM:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{\mu})^2 \quad (20)$$

Porém, como vimos na primeira parte, eles são variáveis aleatórias com distribuições próprias, então podemos utilizar técnicas para saber o quanto bem eles aproximam  $\mu$  e  $\sigma^2$ , mas antes, temos que descobrir suas distribuições amostrais! Acontece que a distribuição de  $\hat{\mu}$  depende de  $\sigma^2$ , porém, veremos que a distribuição conjunta de  $\hat{\mu}$  e  $\hat{\sigma}^2$  nos permite inferir  $\mu$  sem referenciar  $\sigma$

### 3.1 Independência da Média e Variância Amostrais

**Teorema 3.1.1:** Suponha que as variáveis  $X_1, \dots, X_n$  são iid, onde  $X_j \sim N(0, 1)$ . Seja  $Q$  uma matriz ortogonal  $n \times n$  e  $\underline{Y} = Q\underline{X}$  onde  $\underline{X} = (X_1, \dots, X_n)^T$ , então  $Y_j \sim N(0, 1)$  e  $Y_1, \dots, Y_n$  são iid, além de que  $\sum X_i^2 = \sum Y_i^2$

*Demonstração:* A distribuição conjunta de  $X_1, \dots, X_n$  é dada por:

$$f_X(\underline{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n x_i^2 \right\} \quad (21)$$

Sabemos que, se  $Z = h(Y)$  com  $h$  monótona, então

$$f_Z(z) = \frac{f_Y(y)}{|h'(y)|} \quad (22)$$

Se considerarmos  $h(x) = Qx$ , então  $\underline{Y} = h(\underline{X})$ , logo:

$$f_Y(y) = \frac{f_X(Q^T y)}{|\det(Q)|} \quad (23)$$

Porém,  $\det(Q) = 1$  pois  $Q$  é ortogonal. Além disso, sabemos que  $\|x\|^2 = \|Qx\|^2$ , então:

$$f_Y(y) = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n y_i^2 \right\} \quad (24)$$

Logo, chegamos que  $Y_j \sim N(0, 1)$  e são iid □

Agora vamos demonstrar um dos teoremas mais surpreendentes da estatística

**Teorema 3.1.2** (Independência da Média e Variância Amostral): Sejam  $X_1, \dots, X_n \sim N(\mu, \sigma^2)$  e dados os estimadores:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{\mu})^2 \quad (25)$$

As variáveis aleatórias  $\hat{\mu}$  e  $\hat{\sigma}^2$  são **INDEPENDENTES** entre si. Junto desse teorema, também mostraremos que:

$$\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \hat{\mu})^2 \sim \chi_{n-1}^2 \quad (26)$$

*Demonstração:* Primeiro vamos provar considerando que  $\mu = 0$  e  $\sigma^2 = 1$ , de forma que utilizaremos esse resultado para generalizar posteriormente.

Vamos primeiro definir  $u = (\frac{1}{\sqrt{n}} \ \dots \ \frac{1}{\sqrt{n}})^T$ , então construímos uma matriz  $Q$  utilizando Gram-Schmidt de forma que  $u$  seja a primeira linha dessa matriz. Então definimos:

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = Q \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \quad (27)$$

Vimos pelo Teorema 3.1.1 que  $Y_j$  são iid e são normais padrão também. Guarde essa informação! Não é difícil ver que:

$$Y_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i = \hat{\mu}\sqrt{n} \quad (28)$$

Como  $\sum_{i=1}^n X_i^2 = \sum_{i=1}^n Y_i^2$ , então:

$$\sum_{i=2}^n Y_i^2 = \sum_{i=1}^n Y_i^2 - Y_1^2 = \sum_{i=1}^n X_i^2 - n(\hat{\mu})^2 = \sum_{i=1}^n (X_i - \hat{\mu})^2 \quad (29)$$

Ou seja,  $\hat{\mu}$  e  $\hat{\sigma}^2$  são independentes! Dado esse resultado, consideremos agora média e variância não-padrões. Então vamos definir:

$$Z_i = \frac{X_i - \mu}{\sigma} \quad (30)$$

Então  $Z_1, \dots, Z_n$  são iid. Sabemos que  $\bar{Z}_n$  e  $\sum(Z_i - \bar{Z}_n)$  são independentes. Perceba também que, como  $\sum(Z_i - \bar{Z}_n) = \sum_{i=2}^n Y_i$ , então  $\sum(Z_i - \bar{Z}_n) \sim \chi_{n-1}^2$ . Porém, sabemos também que  $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ , então

$$\begin{aligned} \bar{Z}_n &= \frac{\bar{X}_n - \mu}{\sigma} \\ \Rightarrow \sum (Z_i - \bar{Z}_n)^2 &= \frac{1}{\sigma^2} \sum (X_i - \bar{X}_n)^2 \end{aligned} \quad (31)$$

Logo,  $\bar{X}_n$  e  $\frac{1}{n} \sum (X_i - \bar{X}_n)^2$  são independentes e  $\frac{1}{\sigma^2} \sum (X_i - \bar{X}_n)^2 \sim \chi_{n-1}^2$   $\square$

Esse resultado é interessante pois, em certas ocasiões, podemos querer saber a seguinte probabilidade:

$$\mathbb{P}\left(|\hat{\mu} - \mu| \leq \frac{1}{5}\sigma, |\hat{\sigma} - \sigma| \leq \frac{1}{5}\sigma\right) \geq \frac{1}{2} \quad (32)$$

Já que ela indica uma probabilidade de proximidade entre meus estimadores e meus parâmetros. Porém, pelo Teorema 3.1.2, podemos separar essa probabilidade em:

$$\underbrace{\mathbb{P}\left(|\hat{\mu} - \mu| \leq \frac{1}{5}\sigma\right)}_{p_1} \underbrace{\mathbb{P}\left(|\hat{\sigma} - \sigma| \leq \frac{1}{5}\sigma\right)}_{p_2} \geq \frac{1}{2} \quad (33)$$

e isso simplifica bastante nossas contas! Vamos definir  $U \sim N(0, 1)$ , então podemos reescrever as probabilidades como:

$$p_1 = \mathbb{P}\left(\frac{\sqrt{n}}{\sigma} |\hat{\mu} - \mu| < \frac{1}{5}\sqrt{n}\right) = \mathbb{P}\left(|U| < \frac{1}{5}\sqrt{n}\right) \quad (34)$$

definindo  $V = \frac{n}{\sigma^2} \hat{\sigma}^2$ , sabemos que  $V \sim \chi^2_{n-1}$ , então

$$\begin{aligned} p_2 &= \mathbb{P}\left(-\frac{1}{5}\sigma \leq \hat{\sigma} - \sigma \leq \frac{1}{5}\sigma\right) = \mathbb{P}\left(\frac{4}{5}\sigma \leq \hat{\sigma} \leq \frac{6}{5}\sigma\right) \\ &= \mathbb{P}\left(\frac{16}{25}\sigma^2 \leq \hat{\sigma}^2 \leq \frac{36}{25}\sigma^2\right) = \mathbb{P}\left(\frac{16}{25}n \leq V \leq \frac{36}{25}n\right) \end{aligned} \quad (35)$$

e como  $V \sim \chi^2_{n-1}$ , basta consultar uma tabela ou um software para descobrir esses quantis

# **Distribuições $t$**

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim N(0, 1) \quad (36)$$

Porém, podemos não saber  $\sigma$  e queremos substituir, por exemplo, por seu EMV, então qual seria a distribuição de

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\hat{\sigma}} \quad (37)$$

**Definição 4.1** (Distribuição  $t$  com  $m$  graus de liberdade): Se  $Z \sim N(0, 1)$  e  $Y \sim X_m^2$ , então

$$X = \frac{Z}{\sqrt{\frac{Y}{m}}} \sim t_m \quad (38)$$

onde dizemos  $t$  com  $m$  graus de liberdade

**Teorema 4.1** (PDF): A pdf de  $X \sim t_m$  é:

$$f_X(x) = \frac{\Gamma(\frac{m+1}{2})}{(m\pi)^{\frac{1}{2}}\Gamma(\frac{m}{2})} \left(1 + \frac{x^2}{m}\right)^{-(m+1)/2} \quad (39)$$

*Demonstração:* Lembrando:  $Y \sim X_m^2$  e  $Z \sim N(0, 1)$ . Vamos aplicar as transformações! Sabemos que:

$$f_{XW}(x, w) = f_{YZ}(y, z) \mid \frac{\partial(y, z)}{\partial(x, w)} \mid \quad (40)$$

então denotando  $W = Y$ , temos:

$$Z = X \left(\frac{W}{m}\right)^{\frac{1}{2}} \quad Y = W \quad (41)$$

Então vamos ter que:

$$\begin{aligned} \frac{\partial y}{\partial x} &= 0 & \frac{\partial y}{\partial w} &= 1 \\ \frac{\partial z}{\partial x} &= \left(\frac{w}{m}\right)^{\frac{1}{2}} \end{aligned} \quad (42)$$

então vamos ter que

$$\mid \frac{\partial(y, z)}{\partial(x, w)} \mid = -\left(\frac{w}{m}\right)^{\frac{1}{2}} \quad (43)$$

$$\begin{aligned}
f_{XW}(x, w) &= f_{WZ}(w, z) \left( \frac{w}{m} \right)^{\frac{1}{2}} \\
&= f_W(w) f_Z(z) \left( \frac{w}{m} \right)^{\frac{1}{2}} \quad (\text{Independência de } Y \text{ e } Z) \\
&= f_W(w) f_Z \left( x \left( \frac{w}{m} \right)^{\frac{1}{2}} \right) \left( \frac{w}{m} \right)^{\frac{1}{2}} \\
&= \underbrace{\frac{\left(\frac{1}{2}\right)^{\frac{m}{2}}}{\Gamma(\frac{m}{2})} w^{\frac{m}{2}-1} e^{-\frac{1}{2}w}}_{f_W(w)} \underbrace{\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2 w}{2m}}}_{f_Z(z)} \left( \frac{w}{m} \right)^{\frac{1}{2}}
\end{aligned} \tag{44}$$

reescrevendo para ficar algo mais limpo e unir os termos comuns:

$$f_{XW}(x, w) = \frac{\left(\frac{1}{2}\right)^{\frac{m}{2}}}{\Gamma(\frac{m}{2}) \sqrt{2\pi m}} w^{\frac{m-1}{2}} \exp \left\{ -\frac{1}{2} \left( 1 + \frac{x^2}{m} \right) w \right\} \tag{45}$$

Agora, para obtermos a marginal de  $X$ , precisamos integrar isso tudo com relação a  $w$ . Porém, basta integrarmos aquilo que é em função apenas de  $w$ , as constantes nós podemos adicionar novamente depois, então:

$$f_X(x) \propto \int_0^\infty w^{\frac{m-1}{2}} \exp \left\{ -\frac{1}{2} \left( 1 + \frac{x^2}{m} \right) w \right\} dw \tag{46}$$

Se definirmos  $\alpha = \frac{m+1}{2}$  e  $\beta = \frac{1}{2} \left( 1 + \frac{x^2}{m} \right)$ , então a integral equivale a:

$$f_X(x) \propto \int_0^\infty w^{\alpha-1} e^{-\beta w} dw = \frac{\Gamma(\alpha)}{\beta^\alpha} = \frac{\Gamma(\frac{m+1}{2})}{\left(\frac{1}{2} \left[ 1 + \frac{x^2}{m} \right]\right)^{(m+1)/2}} \tag{47}$$

$$\Rightarrow f_X(x) = \frac{\left(\frac{1}{2}\right)^{\frac{m}{2}}}{\Gamma(\frac{m}{2}) \sqrt{2\pi m}} \Gamma\left(\frac{m+1}{2}\right) \left( \frac{1}{2} \left[ 1 + \frac{x^2}{m} \right] \right)^{-(m+1)/2} \tag{48}$$

juntando tudo, temos:

$$f_X(x) = \frac{\Gamma(\frac{m+1}{2})}{\sqrt{m\pi} \Gamma(\frac{m}{2})} \left( 1 + \frac{x^2}{m} \right)^{-(m+1)/2} \tag{49}$$

□

## 4.1 Propriedades

**Teorema 4.1.1:** Se  $T \sim t_m$ , então:

$$\begin{aligned}
\mathbb{E}[T] &= 0 \quad (m > 1) \\
\mathbb{V}[T] &= \frac{m}{m-2} \quad (m > 2)
\end{aligned} \tag{50}$$

*Demonstração:* Seja  $Z \sim N(0, 1)$  e  $W \sim \chi_m^2$ , sabemos que

$$T = \frac{Z}{\sqrt{\frac{W}{m}}} \sim t_m \quad (51)$$

porém, temos que  $T|W = w \sim N(0, \frac{m}{w})$ , logo:

$$\mathbb{E}[T|W = w] = 0 \quad (52)$$

pela lei de adão:

$$\mathbb{E}[\mathbb{E}[T|W = w]] = \mathbb{E}[T] = \mathbb{E}[0] = 0 \quad (53)$$

No mesmo raciocínio, lembrando a lei de EVA

$$\begin{aligned} \mathbb{V}[X] &= \mathbb{E}[\mathbb{V}[X|Y]] + \mathbb{V}[\mathbb{E}[X|Y]] \\ \Rightarrow \mathbb{V}[T] &= \mathbb{E}[\mathbb{V}[T|W]] + \mathbb{V}[\mathbb{E}[T|W]] \end{aligned} \quad (54)$$

sabemos que  $\mathbb{E}[T|W] = 0$ , então basta calcularmos  $\mathbb{V}[T|W]$  que, como vimos antes, vai ser  $\frac{W}{m}$ , então:

$$\mathbb{V}[T] = \mathbb{E}\left[\frac{m}{W}\right] \quad (55)$$

Sabemos que  $W$  é uma  $\Gamma(\frac{m}{2}, \frac{1}{2})$ , então  $W^{-1}$  é uma Gamma Inversa, logo, sua média vai ser:

$$\mathbb{E}[W^{-1}] = \frac{\frac{1}{2}}{\frac{m}{2} - 1} = \frac{\frac{1}{2}}{\frac{m-2}{2}} = \frac{1}{m-2} \quad (56)$$

logo:

$$\mathbb{V}[T] = \frac{m}{m-2} \quad (57)$$

□

**Teorema 4.1.2** (Divergência dos Momentos): Se  $T \sim t_m$ , então  $\mathbb{E}[T^p]$  diverge se  $p \geq m$ . Se  $m$  é inteiro, então apenas os  $m - 1$  primeiros momentos existem

**Teorema 4.1.3:** Sejam  $X_1, \dots, X_n \sim N(\mu, \sigma^2)$  e  $\sigma' = \left(\frac{1}{n-1} \sum (X_i - \bar{X}_n)^2\right)^{\frac{1}{2}}$ , então

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma'} \sim t_{n-1} \quad (58)$$

*Demonstração:* Defina  $S_n^2 = \sum (X_i - \bar{X}_n)^2$ ,  $Z = \sqrt{n}(\bar{X}_n - \mu)/\sigma$  e  $Y = S_n^2/\sigma^2$ . Sabemos que  $Y$  e  $Z$  são independentes e  $Y \sim \chi_{n-1}^2$ . Definimos então:

$$U = \frac{Z}{\sqrt{\frac{Y}{n-1}}} \quad (59)$$

que é uma  $t_{n-1}$  por definição. Porém, perceba que:

$$U = \frac{\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}}{\frac{1}{\sigma}\sqrt{\frac{S_n^2}{n-1}}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma'} \quad (60)$$

□

Essa propriedade é interessante, pois saímos de variáveis que dependiam diretamente de  $\sigma$  para uma variável que tem distribuição que **não depende** de  $\sigma$

**Teorema 4.1.4:** Uma distribuição  $t_1$  é equivalente a uma distribuição de *Cauchy*

**Definição 4.1.1** (Distribuição de Cauchy): Se  $X \sim \text{Cauchy}(x_0, \gamma)$ , então temos:

$$f_X(x) = \frac{1}{\pi\gamma \left[ 1 + \left( \frac{x-x_0}{\gamma} \right)^2 \right]} \quad (61)$$

$$F_X(x) = \frac{1}{\pi} \arctan\left(\frac{x-x_0}{\gamma}\right) + \frac{1}{2} \quad (62)$$

Além do fato que:

$$\mathbb{E}[|X|^k] = \infty \quad \forall k \quad (63)$$

# **Intervalos de Confiança**

Essa é uma parte que costuma gerar bastante confusão! Intervalos de confiança adicionam mais informações a um estimador  $\hat{\theta}$  quando não conhecemos o  $\theta$ , de forma que podemos encontrar um intervalo  $[A, B]$  com alta probabilidade de conter  $\theta$

## 5.1 Intervalo de Confiança para a média de uma Normal

Dada a amostra  $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ , sabemos que  $U = \sqrt{n}(\bar{X} - \mu)/\sigma' \sim t_{n-1}$ . Eu gostaria de achar um intervalo no qual eu tenho uma chance boa de encontrar minha média, então eu gostaria de algo do tipo:

$$\mathbb{P}(-c < U < c) = \gamma \quad (64)$$

O método mais comum é calcular diretamente o  $c$  que torna a equação (64) verdadeira. Isso é equivalente a dizer:

$$\mathbb{P}\left(\bar{X}_n - \frac{c\sigma'}{\sqrt{n}} < \mu < \bar{X}_n + \frac{c\sigma'}{\sqrt{n}}\right) = \gamma \quad (65)$$

Vale ressaltar que essa probabilidade é referente a distribuição conjunta de  $\bar{X}_n$  e  $\sigma'$  para valores **fixos** de  $\mu$  e  $\sigma$  (Independentemente de sabermos eles ou não). Então vamos tentar achar o  $c$  que satisfaz isso

$$\mathbb{P}(-c < U < c) = \gamma \Leftrightarrow T_{n-1}(c) - T_{n-1}(-c) = \gamma \quad (66)$$

Pela simetria de  $t$  em 0, posso reescrever como:

$$2T_{n-1}(c) - 1 = \gamma \Rightarrow c = T_{n-1}^{-1}\left(\frac{1+\gamma}{2}\right) \quad (67)$$

Então, depois que descobrimos  $c$ , nosso intervalo de confiança vira:

$$\begin{aligned} A &= \bar{X}_n - \frac{\sigma'}{\sqrt{n}} T_{n-1}^{-1}\left(\frac{1+\gamma}{2}\right) \\ B &= \bar{X}_n + \frac{\sigma'}{\sqrt{n}} T_{n-1}^{-1}\left(\frac{1+\gamma}{2}\right) \end{aligned} \quad (68)$$

Dada essa noção inicial, vamos definir formalmente esses intervalos:

**Definição 5.1.1** (Intervalo de confiança): Seja  $\underline{X} = (X_1, \dots, X_n)$  uma amostra de uma distribuição indexada pelo(s) parâmetro(s)  $\theta$  e seja  $g(\theta) : \Omega \rightarrow \mathbb{R}$ . Seja também  $A$  e  $B$  duas estatísticas ( $A \leq B$ ) que satisfazem:

$$\mathbb{P}(A < g(\theta) < B) \geq \gamma \quad (69)$$

O intervalo aleatório  $(A, B)$  é chamado de intervalo de confiança  $\gamma$  para  $g(\theta)$  ou de intervalo de confiança  $100\gamma\%$  para  $g(\theta)$ . Depois que  $X_1, \dots, X_n$  foi observado e o intervalo  $A = a$  e  $B = b$  foi computado, chamamos o valor observado do intervalo de **valor observado do intervalo de confiança**. Se a equação (69) vale a igualdade  $\forall c \in (A, B)$ , então chamamos esse intervalo de **exato**

Aqui eu vou definir melhor a interpretação com relação a essa definição, que pode ser um pouco confusa. A interpretação do intervalo  $A, B$  em si é bem direta, representa um intervalo **aleatório** que tem probabilidade  $\gamma$  de conter  $g(\theta)$ . Porém, ao observamos as amostras e calcularmos  $A = a$  e  $B = b$ , o intervalo  $(a, b)$  **não necessariamente contém  $g(\theta)$  com probabilidade  $\gamma$** , como assim? Lembra que  $(A, B)$  é um intervalo **aleatório**, enquanto  $(a, b)$  é uma das muitas possíveis ocorrências desse intervalo! A interpretação correta é, que quanto mais repetimos o experimento e computamos  $(a, b)$  e armazenamos

esses valores observados de intervalo, uma fração  $\gamma$  deles contém  $g(\theta)$ , porém, **não sabemos dizer quais contém e quais não contém**

**Teorema 5.1.1:** Dada a amostra  $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ .  $\forall \gamma \in [0, 1]$ , o intervalo  $(A, B)$  com seguintes pontos:

$$\begin{aligned} A &= \bar{X}_n - \frac{\sigma'}{\sqrt{n}} T_{n-1}^{-1}\left(\frac{1+\gamma}{2}\right) \\ B &= \bar{X}_n + \frac{\sigma'}{\sqrt{n}} T_{n-1}^{-1}\left(\frac{1+\gamma}{2}\right) \end{aligned} \quad (70)$$

é um intervalo de confiança  $\gamma$ -exato

## 5.2 Intervalos de Confiança Unilaterais

Nós vimos como encontrar intervalos aleatórios  $(A, B)$  que tem probabilidade  $\gamma$  de conter o parâmetro  $\theta$ , porém, podem acontecer situações que apenas obter um limite superior ou inferior seja suficiente para nós

Dados  $\gamma_1$  e  $\gamma_2$  com  $\gamma_2 > \gamma_1$  e  $\gamma_2 - \gamma_1 = \gamma$ , então:

$$\mathbb{P}(T_{n-1}^{-1}(\gamma_1) < U < T_{n-1}^{-1}(\gamma_2)) = \gamma \quad (71)$$

E então obtemos que, perante todos os intervalos aleatórios possíveis, o intervalo de confiança  $\gamma$  com o menor tamanho é o simétrico

$$\gamma_1 = 1 - \gamma_2 \quad (72)$$

Porém, há casos que um intervalo não-simétrico é útil (Como mencionei o caso anterior de apenas limites superiores ou intefiores)

**Definição 5.2.1** (Intervalo de Confiança Generalizado): Seja  $\underline{X} = (X_1, \dots, X_n)$  uma amostra de uma distribuição parametrizada por  $\theta$ . Seja  $g(\theta) : \Omega \rightarrow \mathbb{R}$  e seja  $A$  uma estatística tal que:

$$\mathbb{P}(A < g(\theta)) \geq \gamma \quad \forall \theta \quad (73)$$

Então o intervalo aleatório  $(A, +\infty)$  é chamado de intervalo de confiança unilateral  $\gamma$  de limite inferior  $A$ . A mesma definição vale para a estatística  $B$  tal que:

$$\mathbb{P}(g(\theta) < B) \geq \gamma \quad (74)$$

Então o intervalo aleatório  $(-\infty, B)$  é chamado de intervalo de confiança unilateral  $\gamma$  de limite superior  $B$ . Se a desigualdade “ $\geq \gamma$ ” é uma igualdade para todo  $\theta$ , então tanto o intervalo quanto os limites são chamados de exatos

**Teorema 5.2.1** (Intervalo unilateral da média da normal): Seja  $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ , as estatísticas a seguir são, respectivamente, limites inferior e superior exatos com coeficiente  $\gamma$  para  $\mu$ :

$$\begin{aligned} A &= \bar{X}_n - T_{n-1}^{-1}(\gamma) \frac{\sigma'}{\sqrt{n}} \\ B &= \bar{X}_n + T_{n-1}^{-1}(\gamma) \frac{\sigma'}{\sqrt{n}} \end{aligned} \tag{75}$$

### 5.3 Intervalo de confiança para outros parâmetros

Até agora, só vimos a aplicação de intervalos de confiança para a distribuição normal, mas por quê? Pois a normal possui propriedades que tornam encontrar os intervalos de confiança mais fáceis, como por exemplo, encontrarmos estatísticas (Por exemplo  $T = \sqrt{n}(\bar{X}_n - \mu)/\sigma'$ ) que não dependem do parâmetro que queremos estimar, e isso na verdade é uma definição útil que pode nos ajudar:

**Definição 5.3.1** (Pivô): Seja  $\underline{X} = (X_1, \dots, X_n)$  uma amostra de uma distribuição parametrizada por  $\theta$  e  $V(\theta, \underline{X})$  uma variável aleatória tal que sua distribuição **não depende de  $\theta$**  e é a mesma  $\forall \theta$ , então chamamos  $V(\theta, \underline{X})$  de **quantidade pivotal ou pivô**

Podemos então utilizar dessa definição para construir intervalos de confiança. Porém, para isso, precisamos de uma “função inversa” desse  $V$ , algo do tipo:

$$r(V(\theta, \underline{X}), \underline{X}) = g(\theta) \tag{76}$$

**Teorema 5.3.1:** Seja  $\underline{X} = (X_1, \dots, X_n)$  uma amostra de uma distribuição parametrizada por  $\theta$ . Suponha que existe um pivô  $V$ . Seja  $F_V(v)$  a CDF de  $V$  e contínua. Assuma também que a função  $r$  tal qual a equação (76) existe e é estritamente crescente em  $v$  para cada  $\underline{x}$ . Seja  $\gamma \in (0, 1)$  e  $\gamma_2 > \gamma_1$  tal que  $\gamma_2 - \gamma_1 = \gamma$ , então as seguintes estatísticas são endpoints de um intervalo de confiança  $\gamma$ -exato para  $g(\theta)$

$$\begin{aligned} A &= r(F_V^{-1}(\gamma_1), \underline{x}) \\ B &= r(F_V^{-1}(\gamma_2), \underline{x}) \end{aligned} \tag{77}$$

Se  $r$  é estritamente decrescente, então invertemos  $A$  e  $B$

*Demonstração:* Se  $r(\theta, \underline{x})$  é estritamente crescente em  $v$  para todo  $\underline{x}$ , então:

$$V(\theta, \underline{X}) < c \Leftrightarrow g(\theta) < r(c, \underline{X}) \tag{78}$$

Defina  $c = F_V^{-1}(\gamma_i)$  para  $i = 1, 2$ , então obtemos:

$$\begin{aligned} \mathbb{P}(g(\theta) < A) &= \gamma_1 \\ \mathbb{P}(g(\theta) < B) &= \gamma_2 \end{aligned} \tag{79}$$

Como  $V$  tem distribuição contínua e  $r$  é estritamente crescente, então:

$$\mathbb{P}(A = g(\theta)) = \mathbb{P}(V(\theta, \underline{X}) = F_V^{-1}(\gamma_1)) = 0 \tag{80}$$

Similarmente com  $\mathbb{P}(B = g(\theta))$ , então combinamos as duas equações na equação (79) para obter  $\mathbb{P}(A < g(\theta) < B) = \gamma$   $\square$

*Exemplo 5.3.1:* Seja  $X_1, \dots, X_n$  uma amostra aleatória de uma distribuição normal com média  $\mu$  e variância  $\sigma^2$  desconhecidos. Vimos anteriormente que:

$$V(\theta, \underline{X}) = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \sim \chi_{n-1}^2 \quad \forall \theta = (\mu, \sigma^2) \quad (81)$$

Logo,  $V$  é um pivô, de forma que conseguimos utilizá-lo para achar intervalos de confiança para  $\sigma^2$

Porém é bem comum que o pivô não exista em casos discretos

*Exemplo 5.3.2:* Seja  $\theta$  a proporção de sucessos em uma população muito grande de pacientes tratados com imipramina. Suponha que os clínicos desejem uma variável aleatória  $A$  tal que, para todo  $\theta$ , tenhamos

$$\Pr(A < \theta) \geq 0.9 \quad (82)$$

Isto é, eles querem ter **90% de confiança** de que a proporção de sucesso seja **pelo menos  $A$** . Os dados observáveis consistem no número  $X$  de sucessos em uma amostra aleatória de  $n = 40$  pacientes. Nenhuma variável pivotal existe neste exemplo, e os intervalos de confiança são mais difíceis de construir

# Análise Bayesiana de Amostras Normais

Nesse capítulo, vamos fazer uma análise bayesiana completa quando o problema trata de amostras de uma distribuição **Normal**. Para facilitar algumas contas, vamos trocar a definição usual com variância para a definição com **precisão**

**Definição 6.1** (Precisão): A precisão de uma distribuição  $N(\mu, \sigma^2)$  é

$$\tau = \frac{1}{\sigma^2} \quad (83)$$

**Teorema 6.1** (Densidade da Normal): Seja  $X \sim N(\mu, \tau)$ , temos que a **função de densidade probabilística** de  $X$  é:

$$f(x|\mu, \tau) = \left( \frac{\tau}{2\pi} \right)^{\frac{1}{2}} \exp\left( -\frac{\tau}{2}(x - \mu)^2 \right) \quad (84)$$

**Corolário 6.1.1** (*Likelihood*): Sejam  $X_1, \dots, X_n | \mu, \tau \sim N(\mu, \tau)$ , temos que a função de verossimilhança é dada por:

$$f(\underline{x}|\mu, \tau) = \left( \frac{\tau}{2\pi} \right)^{\frac{n}{2}} \exp\left( -\frac{1}{2}\tau \sum_{i=1}^n (x_i - \mu)^2 \right) \quad (85)$$

## 6.1 Família de Conjugados

**Teorema 6.1.1** (Família de Conjugados): Suponha que  $X_1, \dots, X_n | \mu, \tau \sim N(\mu, \tau)$  e temos que  $\mu|\tau \sim N(\mu_0, \lambda_0\tau_0)$  e  $\tau \sim \Gamma(\alpha_0, \beta_0)$ , então a posteriori de  $\mu$  e  $\tau$   $[p(\mu, \tau|\underline{x})]$  é:

$$\begin{aligned} \mu, \tau|\underline{x} &\sim N(\mu_1, \lambda_1\tau) \\ \mu_1 &= \frac{\lambda_0\mu_0 + n\bar{x}_n}{\lambda_0 + n} \quad \lambda_1 = \lambda_0 + n \end{aligned} \quad (86)$$

$$\begin{aligned} \tau &\sim \Gamma(\alpha_1, \beta_1) \\ \alpha_1 &= \alpha_0 + \frac{n}{2} \quad \beta_1 = \beta_0 + \frac{1}{2}s_n^2 + \frac{n\lambda_0(\bar{x}_n - \mu_0)^2}{2(\lambda_0 + n)} \end{aligned} \quad (87)$$

Essa família de conjugados é chamada de NormalGamma com parâmetros  $\alpha_0, \beta_0, \mu_0$  e  $\lambda_0$ , de forma que a posteriori de  $\mu, \tau$  é a NormalGamma com parâmetros  $\alpha_1, \beta_1, \mu_1$  e  $\lambda_1$ . Vale lembrar também que:

$$p(\mu, \tau) \propto p(\mu|\tau)p(\tau) \quad (88)$$

Outro ponto é que  $\mu$  e  $\tau$  **não** são independentes, e mesmo que a gente escolha eles de forma que eles sejam independentes a priori, mesmo após uma única observação, eles já viraram dependentes

## 6.2 Marginais

Nós encontramos as distribuições de  $\mu, \tau, \mu|\tau$  e  $\tau$ , porém, qual seria a marginal de  $\mu$ ?

**Teorema 6.2.1** (Marginal de  $\mu$ ): Suponha que  $\mu, \tau \sim \text{NormalGamma}(\mu_0, \lambda_0, \alpha_0, \beta_0)$ , então:

$$\left( \frac{\lambda_0 \alpha_0}{\beta_0} \right)^{\frac{1}{2}} (\mu - \mu_0) \sim t_{2\alpha_0} \quad (89)$$

*Demonstração:*  $\mu | \tau \sim N(\mu_0, \lambda_0 \tau)$ , então temos que:

$$\mathbb{V}[\mu | \tau] = \frac{1}{\lambda_0 \tau} \Rightarrow (\mu - \mu_0) \cdot (\lambda_0 \tau)^{\frac{1}{2}} \sim N(0, 1) \quad (90)$$

Então seja  $p(\tau)$  a marginal de  $\tau$  e  $p(\mu | \tau)$  a pdf condicional de  $\mu$  em  $\tau$

$$p(z, \tau) = \underbrace{(\lambda_0 \tau)^{-\frac{1}{2}} \cdot p\left(\mu = (\lambda_0 \tau)^{-\frac{1}{2}} z + \mu_0 \mid \tau\right)}_{\Phi(z) \rightarrow \text{pdf da } N(0, 1)} p(\tau) \quad (91)$$

Como eu consigo exprimir  $p(z, \tau)$  como a multiplicação de suas marginais, isso significa que  $z$  e  $\tau$  são **independentes**. Definimos então  $Y = 2\beta_0 \tau \Rightarrow Y \sim \Gamma(\alpha_0, \frac{1}{2}) \sim X_{2\alpha_0}^2$ . Ou seja, vamos ter que:

$$U = \frac{Z}{\left(\frac{Y}{2\alpha_0}\right)^{\frac{1}{2}}} \sim t_{2\alpha_0} = \frac{(\lambda_0 \tau)^{\frac{1}{2}} (\mu - \mu_0)}{\left(\frac{2\beta_0 \tau}{2\alpha_0}\right)^{\frac{1}{2}}} = \left(\frac{\lambda_0 \alpha_0}{\beta_0}\right)^{\frac{1}{2}} (\mu - \mu_0) \quad (92)$$

□

Por conta disso, obtemos o seguinte

**Corolário 6.2.1.1** (*Propriedades da Marginal de  $\mu$* ): Se  $\alpha_0 > \frac{1}{2} \Rightarrow \mathbb{E}[\mu] = \mu_0$ . Se  $\alpha_0 > 1 \Rightarrow \mathbb{V}[\mu] = \frac{\beta_0}{\lambda_0(\alpha_0-1)}$

### 6.3 Distribuições Impróprias

Utilizamos esses parâmetros mais por conveniência do que por qualquer outro motivo (Como uma convicção). Para a posteriori, utilizamos os seguintes hiperparâmetros:

$$\alpha_0 = -\frac{1}{2} \quad \beta_0 = 0 \quad \mu_0 = 0 \quad \lambda_0 = 0 \quad (93)$$

assim, obtemos as seguintes pdf's a priori:

$$p(\mu) = 1 \quad p(\tau) = \frac{1}{2} \tau^{-1} \quad p(\mu, \tau) = \frac{1}{\tau} \quad (94)$$

Dessa forma, a posteriori fica:

$$p(\mu, \tau) \propto \left\{ \tau^{\frac{1}{2}} \exp\left[-\frac{n\pi}{2}(\mu - \bar{x}_n)^2\right] \right\} \tau^{\frac{n-1}{2}-1} \exp\left[-\tau \frac{s_n^2}{2}\right] \quad (95)$$

## **Estimadores não-viezados**

Nosso principal objetivo quando estamos fazendo inferência é fazer um estimador  $\delta(\underline{X})$  de  $g(\theta)$  que a distribuição se concentra bem próximo de  $\theta$ , ou seja, na maior parte do tempo, os valores retornados pelo estimador são próximos do verdadeiro valor de  $g(\theta)$ . É com esse objetivo que criamos estimadores **não-viezados**

**Definição 7.1** (Estimador não-viezado): Um estimador  $\delta(\underline{X})$  é não-viezado para  $g(\theta)$  se  $\mathbb{E}_\theta[\delta(\underline{X})] = g(\theta) \forall \theta$

**Definição 7.2** (Viés): O viés de um estimador  $\delta(\underline{X})$  tem o seu **viés** definido como

$$\text{Bias}(g(\theta)) = \mathbb{E}_\theta[\delta(\underline{X})] - g(\theta) \quad (96)$$

Porém, um estimador **não-viezado** não significa que ele é um **bom** estimador ou sequer um estimador viável para a situação. Por exemplo, um estimador que subestima  $g(\theta)$  em 1000 unidades ou superestima vai ser não-viezado, mas sempre retornará valores ruins de aproximação frequentemente. Então para um estimador ser **bom**, ele necessita ter uma baixa variância

**Teorema 7.1:** Seja  $\delta$  um estimador de variância finita, então  $\mathbb{E}_\theta[(\delta - g(\theta))^2] = \mathbb{V}_\theta[\delta] + (\text{Bias}(\delta, g(\theta)))^2$

## 7.1 Estimador não-viezado da Variância

**Teorema 7.1.1:** Seja  $X_1, \dots, X_n$  uma amostra de uma distribuição indexada por  $\theta$  e  $\mathbb{V}[X_i] = \sigma^2$ , então o seguinte estimador é não-viezado:

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad (97)$$

*Demonstração:* Vamos utilizar do fato que:

$$\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2 + n(\bar{X}_n - \mu)^2 \quad (98)$$

Então vamos ter que (Tirando a esperança nos dois lados da equação mostrada anteriormente):

$$\begin{aligned} \mathbb{E}[\hat{\sigma}_0^2] &= \sigma^2 - \frac{\sigma^2}{n} \\ &= \frac{n-1}{n} \sigma^2 \end{aligned} \quad (99)$$

logo, vamos ter que:

$$\frac{n}{n-1} \mathbb{E}[\hat{\sigma}_0^2] = \mathbb{E}[\hat{\sigma}^2] = \sigma^2 \quad (100)$$

□

Esse estimador citado agora é chamado de **variância amostral** em diversas literaturas (No livro do DeGroot, a variância amostral é o estimador com  $1/n$ )

## 7.2 Limitações

- Muitas vezes os estimadores não-viezados possuem uma variância maior que os estimadores viezados, um bom exemplo é que  $\mathbb{V}[\hat{\sigma}^2] \geq \mathbb{V}[\hat{\sigma}_0^2]$
- Não há garantia que estimadores não-viezados existam em toda situação. Um exemplo é que, se  $X_1, \dots, X_n \sim \text{Bern}(p)$ , não existe estimador **não-viezado** de  $\sqrt{p}$
- Estimadores inapropriados, mesmo sendo não-viezados. Por exemplo, se eu tenho uma sequência de bernoullis, e tentar estimar  $p$  pela quantidade de erros até o primeiro sucesso  $X$  (Geométrica). O estimador não viezado seria:

$$\delta(X) = \begin{cases} 1 & \text{se } X = 0 \\ 0 & \text{se } X = 1 \end{cases} \quad (101)$$

- Os estimadores podem ignorar informações. Se você for medir a votlagem de um sistema com um multímetro, e ele retornar 2.5, então podemos pensar que  $\theta$  é 2.5, mas e se o multímetro arredonda tudo que é maior que 3 para 3? Isso muda completamente a distribuição da informação

# Análise e Teste de Hipóteses

De acordo com Natalia Pasternak, presidente do Instituto Questão de Ciência, a ciência pode ser definida como um processo de investigação que leva em conta experimento e observação da natureza, e a partir deles tirar **conclusões provisórias** de acordo com as evidências. Vamos supor que, ao fazermos um experimento, comprovamos a hipótese que estávamos analisando. Ao escrever o artigo científico, é correto afirmar que a nossa hipótese é verdadeira? **Não!** Pois ela foi comprovada em **nossa** estudo, sob as **nossas** condições, com as **nossas** amostras, mas dizer que nossa hipótese está confirmada/rejeitada apenas com o **nossa** estudo é muito subjetivo, então é aí que entra a estatística com a **Análise e Teste de Hipóteses**

## 8.1 Hipóteses Nula e Alternativa

Nós temos  $\theta \in \Omega$  e vamos particionar o espaço em dois conjuntos disjuntos  $\Omega_0$  e  $\Omega_1$  e queremos testar as duas hipóteses:

$$H_0 : \theta \in \Omega_0 \quad H_1 : \theta \in \Omega_1 \quad (102)$$

**Definição 8.1.1:**  $H_0$  é chamada de **hipótese nula** e  $H_1$  a **hipótese alternativa**. Se decidirmos que  $\theta \in \Omega_1$ , então nós **REJEITAMOS**  $H_0$ , se  $\theta \in \Omega_0$ , nós **NÃO REJEITAMOS**  $H_0$

Ué, porque não falamos que **aceitamos** a hipótese  $H_0$ ? Esse modo de visualizar o teste de hipóteses foi popularizado por Ronald Fisher, Jerzy Neyman e Egon Pearson. Essa visualização de assemelha muito ao sistema jurídico, onde seguimos o princípio da presunção de inocência:

- **Hipótese Nula ( $H_0$ ):** Representa o status quo, a crença estabelecida, o “nenhum efeito” ou a “igualdade”. É a hipótese que se presume verdadeira até que haja evidência estatística suficiente para o contrário. (Ex: “O réu é inocente”)
- **Hipótese Alternativa ( $H_1$ ):** É a afirmação que o pesquisador está tentando encontrar evidências para suportar. (Ex: “O réu é culpado.”)

Ou seja, o teste foca em coletar dados que são **inconsistentes** a  $H_0$ . Vamos tentar esclarecer com um exemplo. Você quer saber se uma nova dieta reduziu o peso médio dos participantes.

- $H_0$ : O peso médio não mudou (o efeito da dieta é zero).-  $H_1$ : O peso médio diminuiu.

Se os dados mostrarem uma grande redução de peso, você rejeita a  $H_0$  e conclui que a dieta funcionou. Se os dados mostrarem apenas uma pequena redução, ou um aumento, você não rejeita a  $H_0$ . Você conclui: “Os dados não fornecem evidência suficiente para dizer que a dieta reduziu o peso.” Você não conclui: “A dieta definitivamente não teve efeito.”

*Exemplo 8.1.1 (Exemplo simples):* Temos uma hipótese principal que é “Correr é diminui/intensifica os sintomas da depressão”, então vamos dividir essa hipótese geral nas duas hipóteses que mencionamos anteriormente

- $H_0$ : Correr não afeta em nada os sintomas da depressão
- $H_1$ : Correr diminui/intensifica os sintomas da depressão

Dividimos assim pois, até o momento, queremos comprovar que correr tem algum efeito nos sintomas da depressão, e enquanto não o comprovarmos, assumimos que a atividade física não faz efeito

**Definição 8.1.2 (Hipótese Simples e Composta):** Se  $\Omega_i$  contém apenas 1 valor de  $\theta$ , então  $H_i$  é simples. Se  $\Omega_i$  contém mais que um valor, então  $H_i$  é composta

Quando a hipótese é simples, a distribuição das observações é bem especificada. Já sob hipóteses compostas, dizemos que eles pertencem a uma classe. Uma hipótese nula simples tem a forma:

$$H_0 : \theta = \theta_0 \quad (103)$$

**Definição 8.1.3** (Hipótese unilateral e multilateral): Seja  $\theta \in R$ , hipóteses nulas unidimensionais são da forma  $H_0 : \theta \leq \theta_0$  ou  $H_0 : \theta \geq \theta_0$ . Já hipóteses nulas simples ( $H_0 : \theta = \theta_0$ ) tem hipóteses multilaterais alternativas ( $H_1 : \theta \neq \theta_0$ )

## 8.2 Região Crítica e Testes Estatísticos

Considere o problema de testar as hipóteses:

$$H_0 : \theta \in \Omega_0 \quad H_1 : \theta \in \Omega_1 \quad (104)$$

Seja  $\underline{X} = [X_1, \dots, X_n]$  uma amostra indexada por  $\theta$  desconhecido e  $S$  o conjunto de **todas as saídas possíveis de  $\underline{X}$** . Um estatístico pode especificar um procedimento de teste particionando  $S$  em dois grupos, onde  $S_1$  contém os valores de  $\underline{X}$  onde  $H_0$  será rejeitada e  $S_0$  os valores que  $H_0$  não é rejeitada

**Definição 8.2.1** (Região Crítica): O conjunto  $S_1$  é chamado de **região crítica**

Na maioria dos problemas,  $S_1$  é definido usando uma estatística  $T = r(\underline{X})$

**Definição 8.2.2** (Estatística de Teste e Região de Rejeição): Seja  $T = r(\underline{X})$  uma estatística e  $R \subset \mathbb{R}$ . Suponha que o procedimento de teste das hipóteses seja de forma “Rejeite  $H_0$  se  $T \in R$ ”, então  $T$  é uma **estatística de teste** e  $R$  é a **região de rejeição**

Se definirmos o teste em termos de  $T$  e  $R$  como na definição, então a região crítica é:

$$S_1 := \{\underline{x} | r(\underline{x}) \in R\} \quad (105)$$

*Exemplo 8.2.1:* Ainda na linha de raciocínio do exemplo da atividade física pro combate na depressão, vamos supor que definimos o procedimento  $\delta$  como:

“Rejeite  $H_0$  (Correr não afeta os sintomas da depressão) se o número de pessoas com os sintomas afetados for maior que um valor  $c$ ”

Então podemos definir a estatística de teste como  $\bar{X}$  e a região de rejeição é  $R \subset \mathbb{R}$  com os valores reais maiores que  $c$ . Logo, a região crítica é dada por:

$$S_1 := \{\underline{x} | \bar{X} \in R\} \quad (106)$$

## 8.3 Função de Poder e Tipos de Erro

Seja  $\delta$  um procedimento de teste como definimos antes

**Definição 8.3.1** (Função de Poder): A função  $\pi(\theta|\delta)$  é chamada de **função de poder**. Se  $S_1$  é a região crítica de  $\delta$ , então:

$$\pi(\theta|\delta) = \mathbb{P}(\underline{X} \in S_1 | \theta) \text{ ou } \mathbb{P}(T \in R | \theta) \quad (107)$$

Ou seja, é a probabilidade de que a minha amostra esteja na região crítica dado os meus parâmetros, ou seja, a probabilidade de que vou rejeitar  $H_0$ . A função de poder especial é aquela que:

$$\begin{aligned}\pi(\theta|\delta) &= 0 & \forall \theta \in \Omega_0 \\ \pi(\theta|\delta) &= 1 & \forall \theta \in \Omega_1\end{aligned}\tag{108}$$

Lembrando: Para cada valor  $\theta \in \Omega_0$ , rejeitar  $H_0$  é uma decisão **incorreta** e o mesmo para cada valor  $\theta \in \Omega_1$  e não rejeitar  $H_0$

**Definição 8.3.2** (Tipos de Erro): A decisão errônea de rejeitar uma hipótese nula **verdadeira** é de **Tipo I** (ou primeira ordem). Uma decisão errônea de **não rejeitar** uma hipótese nula **falsa** é chamada de **Tipo II** (ou segunda ordem)

	Aceitar a hipótese nula	Rejeitar a hipótese nula
Hipótese nula é verdadeira	✓	Erro de Tipo I
Hipótese nula é falsa	Erro de Tipo II	✓

Se  $\theta \in \Omega_0$ ,  $\pi(\theta|\delta)$  é a probabilidade de cometemos um erro de Tipo I, já que ele representa a probabilidade de que a amostra esteja na região crítica (rejeitar  $H_0$ ) e, se  $\theta \in \Omega_1$ ,  $1 - \pi(\theta|\delta)$  é a probabilidade de cometemos um erro de Tipo II. No geral, queremos achar  $\delta$  tal que  $\pi(\theta|\delta)$  seja baixo para  $\theta \in \Omega_0$  e alto para  $\theta \in \Omega_1$ , já que isso representa diminuir a probabilidade de cometer cada um dos erros.

Um método muito usado é escolher  $\alpha_0 \in (0, 1]$  tal que:

$$\pi(\theta|\delta) \leq \alpha_0 \quad \forall \theta \in \Omega_0\tag{109}$$

e depois procurar o teste que **maximiza**  $\pi(\theta|\delta)$  satisfazendo a condição (para  $\theta \in \Omega_1$ )

**Definição 8.3.3** (Tamanho de um Teste): Um teste que satisfaz a equação (109) é chamado de **teste de nível  $\alpha_0$**  e que o teste tem nível de significância  $\alpha_0$ . O tamanho  $\alpha(\delta)$  de um teste  $\delta$  é definido por:

$$\alpha(\delta) = \sup_{\theta \in \Omega_0} \pi(\theta|\delta)\tag{110}$$

Ou seja, o tamanho de um teste é a maior probabilidade de cometemos um erro de **Tipo I** possível (já que fazemos o supremo dentre todos os valores de  $\Omega_0$ ) e um teste ter nível de significância  $\alpha_0$  significa que, independente de qual parâmetro de  $H_0$  seja o verdadeiro da distribuição, a chance de cometemos um erro de **Tipo I** sempre será menor que  $\alpha_0$

**Corolário 7.1.1.1:** Um teste  $\delta$  é de nível  $\alpha_0 \Leftrightarrow \alpha(\delta) \leq \alpha_0$

Se a hipótese nula é simples ( $H_0 : \theta = \theta_0$ ), então  $\alpha(\delta) = \pi(\theta_0|\delta)$

## 8.4 Induzindo um nível de significância

Nós queremos testar:

$$\begin{aligned}H_0 &: \theta \in \Omega_0 \\ H_1 &: \theta \in \Omega_1\end{aligned}\tag{111}$$

Seja  $T$  uma estatística e suponha que vamos rejeitar  $H_0$  se  $T \geq c$ . Vamos supor que queremos que nosso teste tenha um nível específico de significância  $\alpha_0$ . Temos:

$$\pi(\theta|\delta) = \mathbb{P}(T \geq c|\theta) \underset{\substack{\rightarrow \\ \text{Queremos}}}{\sup}_{\theta \in \Omega_0} \mathbb{P}(T \geq c|\theta) \leq \alpha_0\tag{112}$$

perceba que o lado direito é não-crescente em  $c$ , então a desigualdade é satisfeita para altos valores de  $c$ , então devemos fazer  $c$  o menor possível sem que a desigualdade seja desfeita, e também queremos que  $\pi(\theta|\delta)$  seja o maior possível para  $\theta \in \Omega_1$ . Quando  $T$  tem distribuição contínua, costuma ser fácil achar um  $c$  apropriado

## 8.5 p-valor

**Definição 8.5.1** (p-valor): O p-valor é o menor nível  $\alpha_0$  ao qual rejeitariam os dados observados. Também chamamos o p-valor de **nível de significância observado**

É o que, macho? Essa definição está muito objetiva e densa, então simplificando um pouco: p-valor é a **probabilidade** de se obter o padrão de resultados que encontramos no nosso estudo ou resultados mais extremos, **considerando a hipótese nula como verdadeira**. Vamos supor que observamos uma amostra  $\underline{x}$  e não fixamos um valor  $\alpha$  para rejeitarmos  $H_0$ , então nos perguntamos “Qual é o menor nível para o qual esses dados ainda seriam considerados extremos o suficiente para rejeitar  $H_0$ ?”. Então o  $p$ -valor segue uma linha diferente do procedimento de teste estabelecido anteriormente. Original:

- Escolhemos um nível  $\alpha$
- Define-se a **região crítica** associada a esse  $\alpha$
- Calcula-se a estatística de teste
- Verificamos se ela cai na região crítica

Agora nós invertemos a lógica

- Em vez de fixar um  $\alpha$ , vamos perguntar “rejeito ou não?”
- Fixamos os dados
- Para quais valores de  $\alpha$  eu rejeitaria? O menor desses será meu  $p$ -valor

Mas por que essa definição é útil? Usamos isso pois, se eu faço um teste em um nível  $\alpha_0$  e rejeito  $H_0$ , simplesmente dizer que rejeitei  $H_0$  no nível  $\alpha_0$  parece vago. Isso não diz o quanto perto estávamos de tomar a outra decisão.

Um experimentador que rejeita a hipótese nula  $\Leftrightarrow$  o p-valor é no máximo  $\alpha_0$ , está usando um teste de significância  $\alpha_0$

## 8.6 Calculando p-valores

Se nossos testes são da forma “Rejeite  $H_0$  quando  $T \geq c$ ” para uma única estatística de teste, tem um jeito direto de calcular p-valores. Para cada  $t$ , deixe  $\delta_t$  o teste que rejeita  $H_0$  quando  $T \geq t$ . Então o p-valor quando  $T = t$  é observado é o tamanho do teste  $\delta_t$ , ou seja, o p-valor é:

$$\sup_{\theta \in \Omega_0} \pi(\theta|\delta_t) = \sup_{\theta \in \Omega_0} \mathbb{P}(T \geq t|\theta) \quad (113)$$

## 8.7 Equivalência de testes e conjuntos de confiança

Os teoremas a seguir mostram a equivalência de intervalos e conjuntos de confiança (o nome é bem intuitivo). Intuitivamente, um intervalo de confiança é um tipo específico de conjunto de confiança (com um tipo específico de regra)

**Teorema 8.7.1:** Seja  $\underline{X} = [X_1, \dots, X_n]$  uma amostra de uma distribuição indexada por um parâmetro  $\theta$ . Seja  $g(\theta)$  uma função e suponha que para todo possível valor  $g_0$  de  $g(\theta)$ , existe um teste  $\delta_{g_0}$  de nível  $\alpha_0$  da hipótese

$$\begin{aligned} H_{0,g_0} &: g(\theta) = g_0 \\ H_{1,g_0} &: g(\theta) \neq g_0 \end{aligned} \quad (114)$$

Para cada possível valor de  $\underline{x}$  de  $\underline{X}$ , defina:

$$\omega(\underline{x}) = \left\{ g_0 | \delta_{g_0} \text{ não rejeita } H_{0,g_0} \text{ se } \underline{X} = \underline{x} \text{ é visto} \right\} \quad (115)$$

e seja  $\gamma = 1 - \alpha_0$ , então o conjunto aleatório  $w(\underline{X})$  satisfaz

$$\mathbb{P}(g(\theta_0) \in \omega(\underline{X}) | \theta = \theta_0) \geq \gamma \quad \forall \theta_0 \in \Omega \quad (116)$$

*Demonstração:* Seja  $\theta_0 \in \Omega$  um elemento arbitrário e defina  $g_0 = g(\theta_0)$ . Como  $\delta_{g_0}$  é um teste de nível  $\alpha_0$ , sabemos que:

$$\mathbb{P}(\delta_{g_0} \text{ não rejeitar } H_{0,g_0} | \theta = \theta_0) \geq 1 - \alpha_0 = \gamma \quad (117)$$

Para cada  $\underline{x}$ ,  $g(\theta_0) \in \omega(\underline{x}) \Leftrightarrow$  o teste  $\delta_{g_0}$  não rejeitar  $H_{0,g_0}$  quando  $\underline{X} = \underline{x}$  é visto

$$\Rightarrow \mathbb{P}(g(\theta_0) \in \omega(\underline{X}) | \theta = \theta_0) = A \quad (118)$$

□

**Definição 8.7.1** (Conjunto de confiança): Se um conjunto aleatório  $\omega(\underline{X})$  satisfaz

$$\mathbb{P}(g(\theta_0) \in \omega(\underline{X}) | \theta = \theta_0) \geq \gamma \quad \forall \theta_0 \in \Omega \quad (119)$$

então o chamamos de conjunto de confiança com coeficiente  $\gamma$  para  $g(\theta)$ . Se a desigualdade for igualdade, o chamamos de exato

**Teorema 8.7.2:** Seja  $X_1, \dots, X_n$  uma amostra aleatória de uma distribuição indexada por  $\theta$  e  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  e seja  $\omega(\underline{X})$  um conjunto de confiança  $\gamma$  para  $g(\theta)$ . Para cada possível valor  $g_0$  de  $g(\theta)$ , construa o teste  $\delta_{g_0} : \delta_{g_0} \text{ não rejeita } H_{0,g_0} \Leftrightarrow g_0 \in \omega(\underline{X})$ . Então  $\delta_{g_0}$  é um teste de nível  $\alpha_0 = 1 - \gamma$

## **Testes $t$**

Nesse capítulo, vamos abordar um caso específico de teste de hipóteses para distribuições normais com média e variância desconhecida

*Exemplo 9.1:* Um instituto médico quer saber a distribuição de quantos dias um paciente internado em UTI's de hospitais permanece internado. Foram coletadas informações de  $n = 30$  hospitais por todo o estado. Vamos supor que modelamos a quantidade de dias que a pessoa se mantém internada como uma variável normal com média  $\mu$  e variância  $\sigma^2$ . Vamos dizer também que queremos testar as hipóteses:

$$H_0 : \mu \geq 200 \quad H_1 : \mu < 200 \quad (120)$$

que teste seria apropriado de se utilizar? Quais são suas propriedades?

## 9.1 Testando Hipóteses sobre a Média de uma Normal quando a Variância é Desconhecida

Consideremos  $X_1, \dots, X_n$  uma amostra de uma distribuição normal com média  $\mu$  e variância  $\sigma^2$  desconhecidas, e também que trabalhamos com as hipóteses:

$$\begin{aligned} H_0 &: \mu \leq \mu_0 \\ H_1 &: \mu > \mu_0 \end{aligned} \quad (121)$$

O espaço paramétrico  $\Omega$  suprime todo vetor bidimensional  $(\mu, \sigma^2)$  com  $\mu \in (-\infty, \infty)$  e  $\sigma^2 > 0$ . Aqui, definimos a estatística de teste  $U$  como:

$$U = \sqrt{n} \cdot \frac{\bar{X}_n - \mu_0}{\sigma'} \quad (122)$$

onde o teste rejeita  $H_0$  se  $U \geq c$ . Sabemos que a distribuição de  $U$  é uma  $t$  com  $n - 1$  graus de liberdade, por isso os testes que utilizam de  $U$  são chamados de **testes  $t$** . Quando invertemos as hipóteses:

$$\begin{aligned} H_0 &: \mu \geq \mu_0 \\ H_1 &: \mu < \mu_0 \end{aligned} \quad (123)$$

o teste vira da forma “rejeite  $H_0$  quando  $U \leq c$ ”

*Exemplo 9.1.1:* No Exemplo 9.1, se a gente quisesse um teste de tamanho  $\alpha_0$ , a gente poderia usar o teste  $t$  que rejeita  $H_0$  se a estatística  $U$  for menor ou igual a um  $c$  (escolhemos  $c$  de forma a fazer o teste ter tamanho  $\alpha_0$ )

## 9.2 Propriedades dos testes $t$

**Teorema 9.2.1** (Nível e Viés dos testes  $t$ ): Seja  $\underline{X} = (X_1, \dots, X_n)$  uma amostra aleatória de uma distribuição normal  $X \sim N(\mu, \sigma^2)$  e  $U$  ser a estatística definida anteriormente. Seja também  $c$  o  $1 - \alpha_0$  quantil da distribuição  $t$  com  $n - 1$  graus de liberdade. Seja  $\delta$  o procedimento que rejeita  $H_0$  na equação (121) se  $U \geq c$ . A função de poder  $\pi(\mu, \sigma^2 | \delta)$  tem as seguintes propriedades:

- i.  $\pi(\mu, \sigma^2 | \delta) = \alpha_0$  quando  $\mu = \mu_0$
- ii.  $\pi(\mu, \sigma^2 | \delta) < \alpha_0$  quando  $\mu < \mu_0$
- iii.  $\pi(\mu, \sigma^2 | \delta) > \alpha_0$  quando  $\mu > \mu_0$
- iv.  $\pi(\mu, \sigma^2 | \delta) \rightarrow 0$  conforme  $\mu \rightarrow -\infty$
- v.  $\pi(\mu, \sigma^2 | \delta) \rightarrow 1$  conforme  $\mu \rightarrow \infty$

Além disso, o teste  $\delta$  tem tamanho  $\alpha_0$  e é não-viezado

*Demonstração:* Se  $\mu = \mu_0$ , então  $U \sim t_{n-1}$ , portanto:

$$\pi(\mu_0, \sigma^2 | \delta) = \mathbb{P}(U \geq c | \mu_0, \sigma^2) = \alpha_0 \quad (124)$$

lembrando que a função de poder, quando  $\theta \in \Omega_0$  é a probabilidade de rejeitarmos  $H_0$  (Erro de **Tipo I**), então vai ser a probabilidade de  $U \geq c$ . Isso prova (i)

Para provar (ii) e (iii), defina:

$$U^* = \sqrt{n} \cdot \frac{\bar{X}_n - \mu}{\sigma'} \quad W = \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma'} \quad (125)$$

Logo,  $U = U^* - W$ . Primeiramente, assuma que  $\mu < \mu_0$ , então  $W > 0$ , então segue que:

$$\begin{aligned} \pi(\mu, \sigma^2 | \delta) &= \mathbb{P}(U \geq c | \mu, \sigma^2) = \mathbb{P}(U^* - W \geq c | \mu, \sigma^2) \\ &= \mathbb{P}(U^* \geq c + W | \mu, \sigma^2) < \mathbb{P}(U^* \geq c | \mu, \sigma^2) \end{aligned} \quad (126)$$

Como  $U^*$  tem distribuição  $t_{n-1}$ , a última probabilidade da equação é igual a  $\alpha_0$ . Isso prova (ii). Para provar (iii), basta assumir que  $\mu > \mu_0$ , logo  $W < 0$ , então o sinal de **menor que** no final da equação vira um **maior que**. A prova de (iv) e (v) são mais complicadas e não serão abordadas  $\square$

**Corolário 9.2.1.1:** Seja  $\underline{X} = (X_1, \dots, X_n)$  uma amostra aleatória de uma distribuição normal  $X \sim N(\mu, \sigma^2)$  e  $U$  ser a estatística definida anteriormente. Seja também  $c$  o  $1 - \alpha_0$  quantil da distribuição  $t$  com  $n - 1$  graus de liberdade. Seja  $\delta$  o procedimento que rejeita  $H_0$  na equação (123) se  $U \leq c$ . A função de poder  $\pi(\mu, \sigma^2 | \delta)$  tem as seguintes propriedades:

- i.  $\pi(\mu, \sigma^2 | \delta) = \alpha_0$  quando  $\mu = \mu_0$
- ii.  $\pi(\mu, \sigma^2 | \delta) > \alpha_0$  quando  $\mu < \mu_0$
- iii.  $\pi(\mu, \sigma^2 | \delta) < \alpha_0$  quando  $\mu > \mu_0$
- iv.  $\pi(\mu, \sigma^2 | \delta) \rightarrow 1$  conforme  $\mu \rightarrow -\infty$
- v.  $\pi(\mu, \sigma^2 | \delta) \rightarrow 0$  conforme  $\mu \rightarrow \infty$

Além disso, o teste  $\delta$  tem tamanho  $\alpha_0$  e é não-viezado

*Exemplo 9.2.1:* Para o Exemplo 9.1, se quiséssemos um teste de nível de significância  $\alpha_0 = 0.1$ , então pelas propriedades, rejeitariammos  $H_0$  se  $U \leq c$  onde  $c = T_{n-1}^{-1}(0.1)$ .

Calcular  $p$ -valores para os testes  $t$  é bem direto ao ponto!

**Teorema 9.2.2** ( $p$ -valores para testes  $t$ ): Suponha que estamos testando ou as hipóteses da equação (121) ou da (123). Seja  $u$  o valor observado da estatística  $U$  e  $T_{n-1}(\cdot)$  a cdf da distribuição  $t_{n-1}$ . Então o  $p$ -valor para as hipóteses da equação (121) é  $1 - T_{n-1}(u)$  e para as hipóteses da equação (123) é  $T_{n-1}(u)$

*Demonstração:* Seja  $T_{n-1}^{-1}(\cdot)$  a função quantil da  $t_{n-1}$ . Nós rejeitariammos a hipótese na equação (121) em um nível  $\alpha_0$  se, e somente se  $u \geq T_{n-1}^{-1}(1 - \alpha_0)$ , que é equivalente a  $\alpha_0 \geq 1 - T_{n-1}(u)$ . Similarmente, rejeitamos as hipóteses da equação (123) se, e somente se  $u \leq T_{n-1}^{-1}(\alpha_0)$ , que é equivalente a  $\alpha_0 \geq T_{n-1}(u)$   $\square$

*Exemplo 9.2.2 (Tamanho de Fibras):* Suponha que os comprimentos, em milímetros, de fibras metálicas produzidas por um determinado processo tenham distribuição normal, com média desconhecida ( $\mu$ ) e variância desconhecida ( $\sigma^2$ ), e que as seguintes hipóteses devam ser testadas:

$$H_0 : \mu \leq 5.2 \quad H_1 : \mu > 5.2 \quad (127)$$

Suponha que os comprimentos de 15 fibras selecionadas aleatoriamente sejam medidos e que se observe que a média amostral ( $\bar{X}_{15}$ ) é 5.4 e que ( $\sigma' = 0.4226$ ). Com base nessas medições, realizaremos um teste  $t$  ao nível de significância ( $\alpha_0 = 0.05$ ).

Como ( $n = 15$ ) e ( $\mu_0 = 5.2$ ), a estatística ( $U$ ) terá distribuição  $t$  com 14 graus de liberdade quando ( $\mu = 5.2$ ). Verifica-se na tabela da distribuição  $t$  que

$$T_{14}^{-1}(0.95) = 1.761. \quad (128)$$

Assim, a hipótese nula ( $H_0$ ) será rejeitada se ( $U > 1.761$ ). Como o valor numérico de ( $U$ ) é 1,833, a hipótese nula ( $H_0$ ) seria rejeitada ao nível de 0.05.

Com o valor observado ( $u = 1.833$ ) para a estatística ( $U$ ) e ( $n = 15$ ), podemos calcular o **p-valor** para as hipóteses utilizando um software computacional que inclua a função de distribuição acumulada das várias distribuições  $t$ . Em particular, obtemos

$$1 - T_{14}(1.833) = 0.0441. \quad (129)$$

Legal, mas será que conseguimos dizer algo sobre a **função poder** de um teste  $t$ ? Se conseguirmos determinar a distribuição de  $U$ , nós conseguimos sim!

**Definição 9.2.1** (Distribuição  $t$  não-central): Seja  $Y$  e  $W$  variáveis aleatórias independentes onde  $W \sim N(\psi, 1)$  e  $Y \sim X_m^2$ , então a distribuição de:

$$X = \frac{W}{\left(\frac{Y}{m}\right)^{1/2}} \quad (130)$$

é chamada de **distribuição  $t$  não-central com  $m$  graus de liberdade e parâmetro de não-centralidade  $\psi$** . Chamaremos a sua cdf de  $T_m(t|\psi)$  ( $T_m(t|\psi) = \mathbb{P} * (X \leq t)$ )

**Teorema 9.2.3:** Seja  $X_1, \dots, X_n$  uma amostra aleatória de uma distribuição normal com média  $\mu$  e variância  $\sigma^2$ . A distribuição da estatística  $U$  é dada por uma distribuição  $t$  não-central com  $n - 1$  graus de liberdade e parâmetro de não-centralidade  $\psi = \sqrt{n}(\mu - \mu_0)/\sigma$ . Seja  $\delta$  o teste que rejeita  $H_0 : \mu \leq \mu_0$  quando  $U \geq c$ . Então a função de poder de  $\delta$  é  $\pi(\mu, \sigma^2 | \delta) = 1 - T_{n-1}(c|\psi)$ . Seja  $\delta'$  o teste que rejeita  $H_0 : \mu \geq \mu_0$  quando  $U \leq c$ , então a função de poder de  $\delta'$  é  $\pi(\mu, \sigma^2 | \delta') = T_{n-1}(c|\psi)$

### 9.3 Teste $t$ pareado

Em vários experimentos, podemos desejar comparar a mesma variável em condições distintas na mesma amostra, então estaríamos interessados em comparar qual condição possui maior média. Nesses casos é comum fazer a subtração entre os valores de cada condição e tratar como uma variável aleatória normal

**Exemplo 9.3.1 ()**: O **National Transportation Safety Board** coleta dados de testes de colisão referentes à quantidade e à localização dos danos em bonecos (**dummies**) colocados nos carros testados. Em uma série de testes, um boneco foi colocado no banco do motorista e outro no banco do passageiro dianteiro de cada carro. Uma das variáveis medidas foi o grau de lesão na cabeça de cada boneco. Entre outros aspectos, há interesse em saber se, e/ou em que medida, a quantidade de lesão na cabeça difere entre o banco do motorista e o banco do passageiro.

Sejam  $(X_1, \dots, X_n)$  as diferenças entre os logaritmos das medidas de lesão na cabeça do lado do motorista e do lado do passageiro. Podemos modelar  $(X_1, \dots, X_n)$  como uma amostra aleatória de uma distribuição normal com média ( $\mu$ ) e variância ( $\sigma^2$ ). Suponha que desejamos testar a hipótese nula ( $H_0 : \mu \leq 0$ ) contra a alternativa ( $H_1 : \mu > 0$ ), ao nível de significância ( $\alpha_0 = 0.01$ ).

Há  $n = 164$  carros. O teste consiste em rejeitar ( $H_0$ ) se

$$U \geq T_{163}^{-1}(0.99) = 2.35. \quad (131)$$

A média das diferenças é  $\bar{x}_n = 0.2199$ . O valor de  $\sigma'$  é 0.5342. A estatística  $U$  é então igual a 5.271. Esse valor é maior que 2.35, e a hipótese nula seria rejeitada ao nível de 0.01. De fato, o **p-valor** é menor que  $1.0 \cdot 10^{-6}$ .

Suponha também que estamos interessados na função poder sob  $H_1$  do teste de nível 0.01. Suponha que a diferença média entre os logaritmos das lesões na cabeça do lado do motorista e do lado do passageiro seja  $\frac{\sigma}{4}$ . Então, o parâmetro de não centralidade é  $\frac{(164)\frac{1}{2}}{4} = 3.20$

## 9.4 Testando uma alternativa bilateral

*Exemplo 9.4.1:* Vamos retomar o Exemplo 9.2.2, mas agora vamos alterar as hipóteses para:

$$H_0 : \mu = 5.2, \quad H_1 : \mu \neq 5.2 \quad (132)$$

Assumiremos novamente que os comprimentos de 15 fibras são medidos, e que o valor de  $U$ , calculado a partir dos valores observados, é 1,833. Testaremos as hipóteses ao nível de significância  $\alpha_0 = 0.05$ .

Como  $\alpha_0 = 0.05$ , nosso valor crítico será o quantil  $1 - \frac{0.05}{2} = 0.975$  da distribuição  $t$  com 14 graus de liberdade. Pela tabela das distribuições  $t$  deste livro, encontramos

$$T_{14}^{-1}(0.975) = 2.145. \quad (133)$$

Assim, o teste **t** especifica a rejeição de  $H_0$  se  $U \leq -2.145$  ou se  $U \geq 2.145$ . Como  $U = 1.833$ , a hipótese  $H_0$  **não** seria rejeitada.

Os valores numéricos nos exemplos enfatizam a importância de decidir se a hipótese alternativa apropriada em um dado problema é unilateral (**one-sided**) ou bilateral (**two-sided**). Quando as hipóteses do Exemplo 9.2.2 foram testadas ao nível de significância 0.05, a hipótese nula  $H_0$ , de que  $\mu \leq 5.2$ , foi rejeitada. Quando as hipóteses desse exemplo foram testadas ao mesmo nível de significância, utilizando os mesmos dados, a hipótese nula  $H_0$ , de que  $\mu = 5.2$ , não foi rejeitada.

**Teorema 9.4.1** (Função de poder de testes  $t$  bilaterais): A função de poder do teste  $\delta$  que rejeita  $H_0 : \mu = \mu_0$  quando  $|U| \geq c$ , onde  $c = T_{n-1}^{-1}(1 - \alpha_0/2)$  pode ser encontrada utilizando a distribuição  $t$  não-central. Se  $\mu \neq \mu_0$ , então  $U$  tem distribuição  $t$  não-central com  $n - 1$  graus de liberdade e parâmetro de não-centralidade  $\psi = \sqrt{n}(\mu - \mu_0)/\sigma$ . A função de poder é:

$$\pi(\mu, \sigma^2 | \delta) = T_{n-1}(-c|\psi) + 1 - T_{n-1}(c|\psi) \quad (134)$$

**Teorema 9.4.2** (*p*-valores de testes *t* bilaterais): Suponha que estamos testando as hipóteses bilaterais  $H_0 : \mu = \mu_0$ ,  $H_1 : \mu \neq \mu_0$ . Seja  $u$  o valor observado da estatística  $U$  e seja  $T_{n-1}(\cdot)$  a cdf de uma  $t_{n-1}$ . Então o *p*-valor é  $2[1 - T_{n-1}(|u|)]$

*Demonstração:* Deixe  $T_{n-1}^{-1}(\cdot)$  denotar a função quantil da  $t_{n-1}$ . Nós rejeitariamos a hipótese nula no nível  $\alpha_0$  se, e somente se,  $|u| \geq T_{n-1}^{-1}(1 - \alpha_0/2)$  que é equivalente a  $T_{n-1}(|u|) \geq 1 - \alpha_0/2$  que é equivalente a  $\alpha_0 \geq 2[1 - T_{n-1}(|u|)]$   $\square$

# **Comparando as médias de duas Distribuições Normais**

No mundo da estatística, é bem comum ocorrer de querermos comparar duas distribuições, suas médias, propriedades, etc. Quando utilizamos duas distribuições normais, os testes e intervalos de confiança são bem similares com os que aparecem quando consideramos apenas uma distribuição

## 10.1 O $t$ -teste biamostral

Primeiramente, considere o problema em que temos duas amostras de variáveis normais (com mesma variância) e queremos saber qual distribuição tem maior média. Especificamente, assumimos que  $\underline{X} = (X_1, \dots, X_m)$  é uma amostra aleatória de  $m$ , onde  $X \sim N(\mu_X, \sigma^2)$  (com  $\mu_X$  e  $\sigma^2$  desconhecidos) e  $\underline{Y} = (Y_1, \dots, Y_n)$  formam uma amostra independente da primeira de  $n$  observações, onde  $Y \sim N(\mu_Y, \sigma^2)$ . Estamos interessados em testar as hipóteses:

$$H_0 : \mu_X \leq \mu_Y \quad H_1 : \mu_X > \mu_Y \quad (135)$$

para cada procedimento  $\delta$ , vamos deixar que  $\pi(\mu_X, \mu_Y, \sigma^2 | \delta)$  seja a **power function** (Definição 8.3.1) de  $\delta$ . Vamos assumir que  $\sigma^2$  é igual para ambas as distribuições, se esse requisito não fosse apropriado para o cenário analisado (posteriormente citarei um exemplo que esse caso é plausível), os testes  $t$  que vamos derivar nas próximas seções não seriam apropriados.

Pensando em um  $\delta$  intuitivo, se a diferença entre as médias  $(\mu_Y - \mu_X)$  for alta, faz sentido rejeitarmos  $H_0$ , certo?

**Teorema 10.1.1** (Estatística  $t$  para amostras duplas): Assumindo a estrutura descrita nos parágrafos anteriores e definindo:

$$\bar{X} = \frac{1}{m} \sum_{i=1}^m X_i \quad \bar{Y} = \frac{1}{n} \sum_{j=1}^n Y_j \quad (136)$$

$$S_X^2 = \sum_{i=1}^m (X_i - \bar{X})^2 \quad S_Y^2 = \sum_{j=1}^n (Y_j - \bar{Y})^2 \quad (137)$$

defina então o teste estatístico:

$$U = \frac{(m+n-2)^{1/2}(\bar{X} - \bar{Y})}{\left(\frac{1}{m} + \frac{1}{n}\right)^{1/2}(S_X^2 + S_Y^2)^{1/2}} \quad (138)$$

Para todos os valores de  $\theta = (\mu_X, \mu_Y, \sigma^2)$  tais que  $\mu_X = \mu_Y$ , temos então que:

$$U \sim t_{m+n-2} \quad (139)$$

*Demonstração:* Assuma que  $\mu_X = \mu_Y$ . Defina as seguintes variáveis aleatórias:

$$\begin{aligned} Z &= \frac{\bar{X} - \bar{Y}}{\left(\frac{1}{m} + \frac{1}{n}\right)^{1/2}\sigma} \\ W &= \frac{S_X^2 + S_Y^2}{\sigma^2} \end{aligned} \quad (140)$$

Agora podemos representar  $U$  como:

$$U = \frac{Z}{(W/(m+n-2))^{1/2}} \quad (141)$$

perceba que se provarmos que  $Z \sim N(0, 1)$ ,  $W \sim X^2_{m+n-2}$ , e  $Z$  e  $W$  são independentes que o teorema está concluído. Desde o começo estamos assumindo que  $X$  e  $Y$  são independentes dado  $\theta$ . Desse fato, segue que toda função de  $X$  é independente de toda função de  $Y$ , em particular,  $(\bar{X}, S_X^2)$  é independente de  $(\bar{Y}, S_Y^2)$ . Pelo Teorema 3.1.2, sabemos que  $\bar{X}$  e  $S_X^2$  são independentes, assim como  $\bar{Y}$  e  $S_Y^2$ , ou seja, todos  $\bar{X}$ ,  $\bar{Y}$ ,  $S_X^2$  e  $S_Y^2$  são independentes entre si, logo,  $Z$  e  $W$  também são independentes.

Segue também do Teorema 3.1.2 que:

$$\frac{S_X^2}{\sigma^2} \sim X^2_{m-1} \quad \frac{S_Y^2}{\sigma^2} \sim X^2_{n-1} \quad (142)$$

e pelas propriedades da  $X^2$  (Teorema 2.1.3), temos que  $W \sim X^2_{m+n-2}$ . Utilizando das propriedades que vimos no curso de Probabilidade, sabemos que  $\bar{X} - \bar{Y}$  tem média  $\mu_X - \mu_Y = 0$  e variância  $\sigma^2/n + \sigma^2/m$ , logo, segue que  $Z \sim N(0, 1)$   $\square$

Um teste  $t$  biamstral com nível de significância  $\alpha_0$  é o procedimento  $\delta$  que rejeita  $H_0$  se  $U \geq T_{m+n-2}^{-1}(1 - \alpha_0)$ . O próximo teorema estabelece algumas propriedades interessantes sobre a **função de poder** de testes  $t$  biamstrais análogos aos do :

**Teorema 10.1.2** (Nível e Viés de Testes  $t$  biamstrais): Seja  $\delta$  um teste  $t$  biamstral definido antes.

A função de poder  $\pi(\mu_X, \mu_Y, \sigma^2 | \delta)$  tem as seguintes propriedades:

- $\pi(\mu_X, \mu_Y, \sigma^2 | \delta) = \alpha_0$  quando  $\mu_X = \mu_Y$
- $\pi(\mu_X, \mu_Y, \sigma^2 | \delta) < \alpha_0$  quando  $\mu_X < \mu_Y$
- $\pi(\mu_X, \mu_Y, \sigma^2 | \delta) > \alpha_0$  quando  $\mu_X > \mu_Y$
- $\pi(\mu_X, \mu_Y, \sigma^2 | \delta) \rightarrow 0$  conforme  $\mu_X - \mu_Y \rightarrow -\infty$
- $\pi(\mu_X, \mu_Y, \sigma^2 | \delta) \rightarrow 1$  conforme  $\mu_X - \mu_Y \rightarrow \infty$

além disso, o teste  $\delta$  tem tamanho  $\alpha_0$  e é não-viezado

Vale ressaltar que, se as hipóteses forem:

$$H_0 : \mu_X \geq \mu_Y \quad H_1 : \mu_X < \mu_Y \quad (143)$$

o teste  $\delta$  correspondente de tamanho  $\alpha_0$  é **rejeitar  $H_0$  quando  $U \leq -T_{m+n-2}^{-1}(1 - \alpha_0)$** .  $P$ -valores são computados de forma muito parecida da forma como se eles fossem de testes  $t$  uniamostral

**Teorema 10.1.3** ( $p$ -valores de testes  $t$  biamstrais): Suponha que estejamos testando as hipóteses da equação (135) ou (143). Seja  $u$  o valor observado da estatística  $U$  (equação (138)) e seja  $T_{m+n-2}(\cdot)$  a cdf da distribuição  $t$  com  $m + n - 2$  graus de liberdade. Então o  $p$ -valor das hipóteses em (135) é  $1 - T_{m+n-2}(u)$  e o  $p$ -valor das hipóteses em (143) é  $T_{m+n-2}(u)$

## 10.2 Poder do Teste

Para cada parâmetro do vetor  $\theta = (\mu_X, \mu_Y, \sigma^2)$ , a função de poder do teste  $t$  biamstral pode ser computada usando a distribuição  $t$  não-central

**Teorema 10.2.1** (Poder do teste  $t$  biamstral): Seja a estatística  $U$  ser definida como na equação (138), então  $U$  tem distribuição não-central  $t$  com  $m + n - 2$  graus de liberdade e parâmetro de não-centralidade

$$\psi = \frac{\mu_X - \mu_Y}{\sigma \left( \frac{1}{m} + \frac{1}{n} \right)^{1/2}} \quad (144)$$

### 10.3 Alternativas Bilaterais

Podemos adaptar os testes  $t$  biamstrais para o caso de hipóteses bilaterais:

$$H_0 : \mu_X = \mu_Y \quad H_1 : \mu_X \neq \mu_Y \quad (145)$$

O teste  $t$  bilateral de tamanho  $\alpha_0$  rejeita  $H_0$  se  $|U| \geq c$  onde  $c = T_{m+n-2}^{-1}(1 - \alpha_0/2)$  e a estatística  $U$  é a mesma definida anteriormente. O  $p$ -valor quando  $U = u$  é observado é igual a  $2[1 - T_{m+n-2}(|u|)]$