

Material Auxiliar - Unidade 5

Intervalos de Confiança e Testes de Hipóteses

Curso de Inferência Estatística

2025

Sumário

I Intervalos de Confiança	3
1 Conceitos Fundamentais	3
1.1 Definição de Probabilidade de Cobertura	3
1.2 Exemplo 1: Cálculo de Probabilidades de Cobertura	3
2 Abordagem por Inversão de Teste de Hipótese	4
2.1 Inversão de um Procedimento de Teste	4
2.2 Exemplo 2: Intervalo Unilateral para Média Normal	5
2.3 Exemplo 3: Intervalo para Parâmetro de Distribuição Exponencial	5
3 Abordagem por Quantidade Pivotal	7
3.1 Definição de Pivô	7
3.2 Exemplo 4: Intervalo para Parâmetro de Distribuição Exponencial (Método Pivotal)	7
3.3 Exemplo 5: Intervalo para Parâmetro de Distribuição Uniforme	8
4 Intervalos de Confiança para Distribuição Normal - Uma Amostra	9
4.1 Exemplo 6: Média com Variância Conhecida	9
4.2 Exemplo 7: Média com Variância Desconhecida	10
4.3 Exemplo 8: Variância e Desvio Padrão	10
4.4 Tabela Comparativa: Intervalos para População Normal - Uma Amostra	11
5 Intervalos de Confiança para Duas Amostras	11
5.1 Abordagem Geral para Duas Amostras	12
5.2 Exemplo 9: Diferença de Médias (Variâncias Iguais)	12
5.3 Exemplo 10: Razão de Variâncias	13
II Teste de Razão de Verossimilhanças (TRV)	14
6 Fundamentos do TRV	15
6.1 Construção	15
6.2 Notas Importantes	16

7 TRV para Uma Amostra - População Normal	16
7.1 Caso 1: Média com Variância Conhecida (Teste Z)	16
7.2 Caso 2: Média com Variância Desconhecida (Teste t)	18
7.3 Caso 3: Variância com Média Conhecida	19
7.4 Caso 4: Variância com Média Desconhecida	21
8 TRV para Duas Amostras	22
8.1 TRV para Comparar Médias	22
8.2 TRV para Comparar Variâncias	25
8.2.1 TRV para $H_0 : \sigma_x^2 = \sigma_y^2$	25
9 Testes Assintóticos: Wald e Escore	27
9.1 Preliminares: Propriedades Assintóticas do EMV	27
9.2 Teste de Wald	28
9.2.1 Construção do Teste de Wald	28
9.2.2 Distribuição Assintótica da Estatística de Wald	28
9.2.3 Regra de Decisão	29
9.2.4 Exemplo: Teste de Wald para Média de População Normal	29
9.3 Teste de Escore (Rao)	29
9.3.1 Função Escore	29
9.3.2 Construção do Teste de Escore	29
9.3.3 Distribuição Assintótica da Estatística de Escore	30
9.3.4 Regra de Decisão	30
9.3.5 Exemplo: Teste de Escore para Média de População Normal	30
9.4 Relação entre TRV, Wald e Escore	31
9.5 Teste de Wald para Hipóteses Parciais	31
10 Teste de Wilcoxon	31
10.1 Teste de Wilcoxon para Uma Amostra (Postos Sinalizados)	31
10.1.1 Hipóteses e Suposições	32
10.1.2 Construção da Estatística	32
10.1.3 Distribuição da Estatística sob H_0	32
10.1.4 Distribuição Assintótica	33
10.1.5 Regra de Decisão	33
10.1.6 Exemplo Ilustrativo	33
10.2 Teste de Wilcoxon para Duas Amostras (Mann-Whitney-Wilcoxon)	34
10.2.1 Hipóteses	34
10.2.2 Construção da Estatística	34
10.2.3 Distribuição da Estatística sob H_0	34
10.2.4 Distribuição Assintótica	35
10.2.5 Regra de Decisão	35
10.2.6 Tratamento de Empates	35
10.2.7 Exemplo Ilustrativo	36
10.3 Vantagens e Desvantagens dos Testes Não-Paramétricos	36
11 Resumo e Observações Finais	36
11.1 Principais Conceitos Abordados	37
11.2 Observações Importantes	37

Parte I

Intervalos de Confiança

1 Conceitos Fundamentais

Vamos começar com o importante conceito de *probabilidade de cobertura*.

1.1 Definição de Probabilidade de Cobertura

Sejam $T_L(\hat{X})$ e $T_U(\hat{X})$ duas estatísticas baseadas numa a.a. $\hat{X} = (X_1, \dots, X_n)$, a probabilidade de cobertura do intervalo aleatório

$$J = [T_L(\hat{X}), T_U(\hat{X})] \quad (1)$$

para o parâmetro desconhecido $\theta \in \Theta \subset \mathbb{R}$ é dada por:

$$P_\theta (\theta \in [T_L(\hat{X}), T_U(\hat{X})]) \quad (2)$$

Na verdade, o *coeficiente de confiança* de J é dado por:

$$\inf_{\theta \in \Theta} \left\{ P_\theta (\theta \in [T_L(\hat{X}), T_U(\hat{X})]) \right\} \quad (3)$$

Na maioria das aplicações, a probabilidade de cobertura não depende do parâmetro e será equivalente ao coeficiente de confiança.

Interpretação: A probabilidade de cobertura representa a probabilidade de que o intervalo aleatório contenha o verdadeiro valor do parâmetro θ . O coeficiente de confiança é o menor valor dessa probabilidade sobre todo o espaço paramétrico, garantindo um nível mínimo de confiança.

1.2 Exemplo 1: Cálculo de Probabilidades de Cobertura

Exemplo 1: Sejam

$$J_1 = (x_1 - 1,96; x_1 + 1,96) \quad \text{e} \quad J_2 = \left(\bar{x} - \frac{1,96}{\sqrt{2}}, \bar{x} + \frac{1,96}{\sqrt{2}} \right)$$

dois intervalos aleatórios tais que $x_1, x_2 \sim N(\mu, 1)$ e

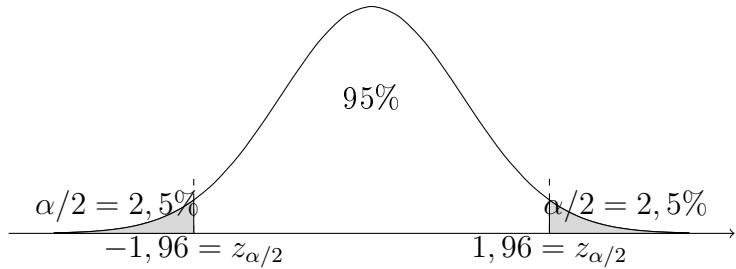
$$\bar{x} = \frac{x_1 + x_2}{2}.$$

Encontre as probabilidades de cobertura de J_1 e J_2 .

Solução: Temos que

$$\begin{aligned} P_\mu(\mu \in J_1) &= P_\mu(\mu \in (x_1 - 1,96; x_1 + 1,96)) \\ &= P_\mu(x_1 - 1,96 < \mu < x_1 + 1,96) \\ &= P_\mu([(x_1 - \mu) < 1,96] \cap [(x_1 - \mu) > -1,96]) \\ &= P_\mu(|x_1 - \mu| < 1,96) \\ &= P_\mu(|Z| < 1,96), \quad Z = x_1 - \mu \sim N(0, 1) \\ &= 95\%. \end{aligned}$$

Explicação: Note que $x_1 - \mu \sim N(0, 1)$ porque $x_1 \sim N(\mu, 1)$. A probabilidade $P(|Z| < 1,96)$ corresponde à área central da distribuição normal padrão, que é exatamente 95%.



Da mesma forma, para J_2 :

$$P_\mu(\mu \in J_2) = P_\mu \left(\mu \in \left(\bar{x} - \frac{1,96}{\sqrt{2}}, \bar{x} + \frac{1,96}{\sqrt{2}} \right) \right)$$

Note que $\bar{x} = \frac{x_1+x_2}{2} \sim N\left(\mu, \frac{1}{2}\right)$, pois a variância da média de duas observações independentes é $\frac{\sigma^2}{n} = \frac{1}{2}$. Portanto,

$$\frac{\bar{x} - \mu}{\sqrt{1/2}} = \sqrt{2}(\bar{x} - \mu) \sim N(0, 1)$$

Assim,

$$P_\mu(\mu \in J_2) = P_\mu \left(|\sqrt{2}(\bar{x} - \mu)| < 1,96 \right) = P(|Z| < 1,96) = 95\%$$

Observação importante: Ambos os intervalos têm a mesma probabilidade de cobertura (95%), mas J_2 é mais eficiente pois utiliza informação de ambas as observações, resultando em um intervalo mais estreito quando comparado a J_1 que usa apenas x_1 .

2 Abordagem por Inversão de Teste de Hipótese

Para construção de intervalos de confiança, podem-se utilizar duas abordagens: (i) inversão do procedimento de teste de hipótese e (ii) usando quantidade pivotal.

2.1 Inversão de um Procedimento de Teste

Em teste de hipótese, a região de não rejeição de H_0 foi denotada como

$$R_C^C = \begin{cases} \{x \in \mathcal{X}^n; T(x|\theta) \leq k\}^C & \text{para } H_1 : \theta > \theta_0, \\ \{x \in \mathcal{X}^n; T(x|\theta) \geq k\}^C & \text{para } H_1 : \theta < \theta_0, \\ (\text{como uma solução plausível}) & \\ \{x \in \mathcal{X}^n; |T(x|\theta)| \leq k\}^C & \text{para } H_1 : \theta \neq \theta_0. \end{cases} \quad (4)$$

O intervalo de confiança é bastante relacionado com R_C . A ideia central é que, se um valor θ_0 não é rejeitado pelo teste, então θ_0 pertence ao intervalo de confiança. Formalmente, o intervalo de confiança é o conjunto de todos os valores de θ que não seriam rejeitados pelo teste.

2.2 Exemplo 2: Intervalo Unilateral para Média Normal

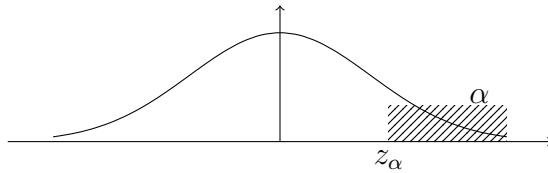
Exemplo 2: Seja X_1, \dots, X_n uma a.a. de $X_i \sim N(\mu, \sigma^2)$ para média desconhecida e $\sigma > 0$ conhecida. Considere que X obedece tanto $H_0 : \mu = \mu_0$ quanto $H_1 : \mu > \mu_0$. Encontre o estimador intervalar para μ de confiança $1 - \alpha$ para $\alpha \in (0, 1)$ pré-fixado.

Solução: Já foi discutido que o teste UMP para $H_0 < H_1$ de nível α tem função crítica:

$$\Psi(x) = \begin{cases} 1, & \sqrt{n} \frac{\bar{X}_n - \mu_0}{\sigma} > z_\alpha, \\ 0, & \text{c.c.} \end{cases} \quad (5)$$

em que $\bar{X}_n = n^{-1} \sum_{i=1}^n x_i$ e $x = (x_1, \dots, x_n)$ é uma a.a. A região de não rejeição é dada por:

$$R_c = \left\{ x \in \mathbb{R}^n : \sqrt{n} \frac{\bar{X}_n - \mu_0}{\sigma} \leq z_\alpha \right\} \quad (6)$$



Note que:

$$P_{\mu_0} \left(\sqrt{n} \frac{\bar{X}_n - \mu_0}{\sigma} \leq z_\alpha \right) = 1 - \alpha \quad (7)$$

Isolando μ_0 na desigualdade, obtemos:

$$\sqrt{n} \frac{\bar{X}_n - \mu_0}{\sigma} \leq z_\alpha \Leftrightarrow \bar{X}_n - \mu_0 \leq z_\alpha \frac{\sigma}{\sqrt{n}} \Leftrightarrow \mu_0 \geq \bar{X}_n - z_\alpha \frac{\sigma}{\sqrt{n}} \quad (8)$$

Portanto,

$$P_{\mu_0} \left(\bar{X}_n - z_\alpha \frac{\sigma}{\sqrt{n}} \leq \mu_0 \right) = 1 - \alpha \quad (9)$$

Daí,

$$P_\mu \left(\mu \geq \bar{X}_n - z_\alpha \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha, \quad \forall \mu \in \mathbb{R}. \quad (10)$$

Isto é,

$$\text{i.c.}_{1-\alpha}(\mu) = \left(\bar{X}_n - z_\alpha \frac{\sigma}{\sqrt{n}}, \infty \right). \quad (11)$$

Interpretação: Este é um intervalo de confiança unilateral inferior. Com probabilidade $1 - \alpha$, o verdadeiro valor de μ é maior ou igual ao limite inferior do intervalo. Note que o limite superior é $+\infty$, o que é característico de intervalos unilaterais.

2.3 Exemplo 3: Intervalo para Parâmetro de Distribuição Exponencial

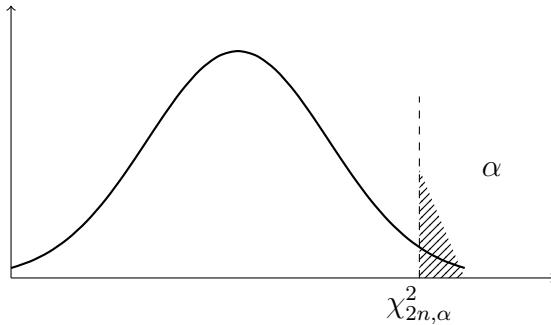
Exemplo 3: Sejam X_1, \dots, X_n uma a.c. de $X \sim \text{Exp}(\theta)$, para $\theta > 0$ desconhecido. Considere que se deseja testar $H_0 : \theta = \theta_0$ vs $H_1 : \theta > \theta_0$. Encontrar o estimador relacionado para o intervalo de confiança de $1 - \alpha$.

Solução: Já foi discutido que o teste UMP para H_0 e H_1 de nível α tem função crítica

$$\psi(x) = \begin{cases} 1, & \frac{2}{\theta_0} \sum_{i=1}^n x_i > \chi_{2n,\alpha}^2 \\ 0, & \frac{2}{\theta_0} \sum_{i=1}^n x_i < \chi_{2n,\alpha}^2 \end{cases} \quad (12)$$

A região de não rejeição é dada por: para $\alpha \in \mathbb{R}^n$,

$$R_c = \left\{ x \in \mathbb{R}^n : \frac{2}{\theta_0} \sum_{i=1}^n x_i < \chi_{2n,\alpha}^2 \right\} \quad (13)$$



Note que:

$$P_\theta \left(\frac{2}{\theta} \sum_{i=1}^n X_i < \chi_{2n,\alpha}^2 \right) = 1 - \alpha \quad (14)$$

Justificativa: A estatística $\frac{2}{\theta} \sum_{i=1}^n X_i$ segue distribuição qui-quadrado com $2n$ graus de liberdade quando $X_i \sim \text{Exp}(\theta)$. Isso decorre do fato de que $2X_i/\theta \sim \chi_2^2$ e a soma de n variáveis qui-quadrado independentes com 2 graus de liberdade cada resulta em χ_{2n}^2 .

Manipulando a desigualdade:

$$\frac{2}{\theta} \sum_{i=1}^n X_i < \chi_{2n,\alpha}^2 \Leftrightarrow \frac{2 \sum_{i=1}^n X_i}{\chi_{2n,\alpha}^2} < \theta \quad (15)$$

Portanto,

$$P_\theta \left(\theta > \frac{2 \sum_{i=1}^n X_i}{\chi_{2n,\alpha}^2} \right) = 1 - \alpha, \quad \forall \theta \in \mathbb{R}^+. \quad (16)$$

Isto é:

$$IC_{1-\alpha}(\theta) = \left(\frac{2 \sum_{i=1}^n X_i}{\chi_{2n,\alpha}^2}, \infty \right). \quad (17)$$

Observação: Se o contraste fosse $H_0 : \theta = \theta_0$ e $H_1 : \theta < \theta_0$, teríamos como região de não rejeição:

$$\mathcal{R} = \left\{ x \in \mathbb{R}^n : \frac{2}{\theta_0} \sum_{i=1}^n x_i > \chi_{2n,1-\alpha}^2 \right\} \quad (18)$$

E o intervalo de confiança seria:

$$P_\theta \left(\theta < \frac{2 \sum_{i=1}^n X_i}{\chi_{2n,1-\alpha}^2} \right) = 1 - \alpha \quad (19)$$

$$IC_{1-\alpha}(\theta) = \left(0, \frac{2 \sum_{i=1}^n X_i}{\chi_{2n,1-\alpha}^2} \right). \quad (20)$$

3 Abordagem por Quantidade Pivotal

3.1 Definição de Pivô

Definição 1: (Pivô) Seja $T(X)$ uma estatística suficiente (mínimal) para θ . Um pivô é uma v.a. U que dependa de T e θ cuja distribuição não dependa de θ .

Observação: No caso da família de locação em $a(\theta)$, a distribuição $\{T - a(\theta)\}$ não depende de θ . No caso de família de escala em $b(\theta)$, a distribuição $\left\{\frac{T}{b(\theta)}\right\}$ não depende de θ . No caso de família de locação e escala em $[a(\theta), b(\theta)]$, a distribuição de $\left\{\frac{T-a(\theta)}{b(\theta)}\right\}$ não depende de θ .

Interpretação: A ideia de um pivô é encontrar uma transformação da estatística e do parâmetro que tenha uma distribuição conhecida e que não dependa do parâmetro desconhecido. Isso permite construir intervalos de confiança de forma direta.

3.2 Exemplo 4: Intervalo para Parâmetro de Distribuição Exponencial (Método Pivotal)

Exemplo 4: Seja $X \sim \text{Exp}(\theta)$ com densidade

$$f(x | \theta) = \frac{1}{\theta} e^{-x/\theta} I_{(0,\infty)}(x). \quad (21)$$

Encontrar um intervalo de confiança $1 - \alpha$ bilateral para θ .

Solução: Note que $U := \frac{X}{\theta}$ tem densidade

$$f_U(u) = \frac{dF_U(u)}{du} = \frac{d\mathbb{P}(X \leq u\theta)}{du} = \theta \cdot \frac{1}{\theta} e^{-u} I_{(0,\infty)}(u). \quad (22)$$

Logo:

$$f_U(u) = e^{-u} I_{(0,\infty)}(u). \quad (23)$$

Portanto U pode ser entendido como um pivô, pois sua distribuição (exponencial padrão com parâmetro 1) não depende de θ .

Note que é possível definir $a, b \in \mathbb{R}$ com $a < b$ tais que

$$\mathbb{P}(U < a) = \mathbb{P}(U > b) = \frac{\alpha}{2}. \quad (24)$$

E, portanto:

$$\mathbb{P}(a < U < b) = 1 - \alpha. \quad (25)$$

Com $\alpha \in (0, 1)$ fixado:

$$\int_0^a e^{-u} du = 1 - e^{-a} = \frac{\alpha}{2} \Rightarrow e^{-a} = 1 - \frac{\alpha}{2} \Rightarrow a = -\log\left(1 - \frac{\alpha}{2}\right). \quad (26)$$

$$\int_b^\infty e^{-u} du = e^{-b} = \frac{\alpha}{2} \Rightarrow b = -\log\left(\frac{\alpha}{2}\right). \quad (27)$$

Dado:

$$P_\theta(a < U < b) = P_\theta\left(a < \frac{X}{\theta} < b\right) = P_\theta\left(\frac{1}{b} < \frac{\theta}{X} < \frac{1}{a}\right) = P_\theta\left(\frac{X}{b} < \theta < \frac{X}{a}\right) = 1 - \alpha \quad (28)$$

Isto é:

$$ic_{1-\alpha}(\theta) = \left(\frac{X}{b}, \frac{X}{a} \right) = \left(\frac{X}{-\log(\alpha/2)}, \frac{X}{-\log(1-\alpha/2)} \right) \quad (29)$$

Observação: Note que $a < b$ implica $\frac{1}{b} < \frac{1}{a}$, garantindo que o intervalo está bem definido.

3.3 Exemplo 5: Intervalo para Parâmetro de Distribuição Uniforme

Exemplo 5: Sejam X_1, \dots, X_n uma a.g. de $X \sim U(0, \theta)$ para θ desconhecido. Encontre o estimador intervalo bilateral para θ com confiança de $1 - \alpha$.

Solução: A estatística $T(X) = X_{\max}$ é suficiente mínima para θ com densidade

$$f_T(t) = \frac{dF_T(t)}{dt} = \frac{d(F_X(t))^n}{dt} = n \left(\frac{t}{\theta} \right)^{n-1} \frac{1}{\theta} \quad (30)$$

Logo:

$$f_T(t) = \frac{n}{\theta^n} t^{n-1} \quad , \quad t \in (0, \theta) \quad (31)$$

Note que $U = \frac{T}{\theta}$ tem densidade

$$f_U(u) = \frac{dF_U(u)}{du} = \frac{dP\left(\frac{T}{\theta} \leq u\right)}{du} = \theta \cdot f_T(u\theta) = \theta \cdot nu^{n-1} \quad (32)$$

$$f_U(u) = nu^{n-1} \mathbb{I}_{(0,1)}(u)$$

que independe de θ , portanto U é um pivô. Pode-se determinar a e b com $0 < a < b < 1$ tais que

$$P(U < a) = P(U > b) = \frac{\alpha}{2}$$

e, portanto,

$$P(a < U < b) = 1 - \alpha$$

Temos que:

$$P(U < a) = \int_0^a nu^{n-1} du = u^n \Big|_0^a = a^n = \frac{\alpha}{2} \quad \Rightarrow \quad a = \left(\frac{\alpha}{2} \right)^{1/n}$$

e

$$P(U > b) = \int_b^1 nu^{n-1} du = u^n \Big|_b^1 = 1 - b^n = \frac{\alpha}{2} \quad \Rightarrow \quad b = \left(1 - \frac{\alpha}{2} \right)^{1/n}$$

Note que

$$\begin{aligned} P(a < U < b) &= 1 - \alpha \quad \Rightarrow \quad P\left(a < \frac{T}{\theta} < b\right) = 1 - \alpha \\ &\Rightarrow \quad P\left(\frac{1}{b} < \frac{\theta}{T} < \frac{1}{a}\right) = 1 - \alpha \end{aligned}$$

$$\Rightarrow P\left(\frac{T}{b} < \theta < \frac{T}{a}\right) = 1 - \alpha$$

e, portanto,

$$IC_{1-\alpha}(\theta) = \left(\frac{T}{b}, \frac{T}{a}\right) = \left(\frac{X_{\max}}{(1-\alpha/2)^{1/n}}, \frac{X_{\max}}{(\alpha/2)^{1/n}}\right)$$

Observação: Note que como $a < b$, temos $\frac{1}{b} < \frac{1}{a}$, garantindo que o limite inferior é menor que o limite superior.

4 Intervalos de Confiança para Distribuição Normal - Uma Amostra

Nesta seção, apresentamos os intervalos de confiança mais comuns para parâmetros de uma população normal, considerando diferentes cenários sobre o conhecimento dos parâmetros.

4.1 Exemplo 6: Média com Variância Conhecida

Exemplo 6: Sejam X_1, \dots, X_n uma amostra de $X \sim N(\mu, \sigma^2)$ com média desconhecida e σ^2 conhecido. Encontrar o intervalo de confiança $1 - \alpha$ bilateral para μ .

Solução: De discussão anterior $T = \bar{X}$ é uma estatística suficiente mínima para μ e

$$T \sim N\left(\mu, \frac{\sigma^2}{n}\right). \quad (33)$$

Aqui, X 's pertencem à família de locação. Note que

$$U = \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \sim N(0, 1) \quad (34)$$

é um pivô. Para $z_{\alpha/2} > 0$ tal que $P(U > z_{\alpha/2}) = \frac{\alpha}{2}$, temos que:

$$P(-z_{\alpha/2} < U < z_{\alpha/2}) = 1 - \alpha \Rightarrow P\left(-z_{\alpha/2} < \sqrt{n} \frac{\bar{X} - \mu}{\sigma} < z_{\alpha/2}\right) = 1 - \alpha \quad (35)$$

ou seja,

$$P\left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha. \quad (36)$$

Portanto,

$$IC_{1-\alpha}(\mu) = \left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right). \quad (37)$$

Interpretação: Este é o intervalo de confiança clássico para a média de uma população normal quando a variância é conhecida. O pivô utilizado é a padronização da média amostral, que segue distribuição normal padrão.

4.2 Exemplo 7: Média com Variância Desconhecida

Exemplo 7: Sejam x_1, \dots, x_n uma a.a. de $X \sim N(\mu, \sigma^2)$ com μ desconhecido e $\sigma \in \mathbb{R}^+$ desconhecido. Encontre o intervalo bilateral de confiança $1 - \alpha$ para μ .

Solução: De discussão anterior, (\bar{X}_n, S_n) é uma estatística conjuntamente suficiente mínima para (μ, σ) . Aqui, X 's pertencem à família de locação e escala.

Considere

$$U = \sqrt{n} \left(\frac{\bar{X}_n - \mu}{S_n} \right) = \frac{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{n-1}{n-1} \cdot \frac{S_n^2}{\sigma^2}}} \sim t_{n-1}, \quad (38)$$

que é um pivô.

Justificativa: O numerador $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$ e o denominador $\sqrt{\frac{S_n^2}{\sigma^2}} = \sqrt{\frac{(n-1)S_n^2/\sigma^2}{n-1}}$, onde $(n-1)S_n^2/\sigma^2 \sim \chi^2_{n-1}$. A razão entre uma normal padrão e a raiz quadrada de uma qui-quadrado dividida por seus graus de liberdade segue distribuição t de Student.

Para $t_{n-1, \alpha/2} > 0$ tal que

$$P(U > t_{n-1, \alpha/2}) = \frac{\alpha}{2}, \quad (39)$$

temos que

$$P(-t_{n-1, \alpha/2} < U < t_{n-1, \alpha/2}) = 1 - \alpha \Rightarrow \quad (40)$$

$$P\left(-t_{n-1, \alpha/2} < \sqrt{n} \frac{\bar{X}_n - \mu}{S_n} < t_{n-1, \alpha/2}\right) = 1 - \alpha \Rightarrow \quad (41)$$

$$P\left(\bar{X}_n - t_{n-1, \alpha/2} \frac{S_n}{\sqrt{n}} < \mu < \bar{X}_n + t_{n-1, \alpha/2} \frac{S_n}{\sqrt{n}}\right) = 1 - \alpha \quad (42)$$

$$\therefore IC_{1-\alpha}(\mu) = \left\{ \bar{X}_n \pm t_{n-1, \alpha/2} \frac{S_n}{\sqrt{n}} \right\}$$

Observação: Quando a variância é desconhecida, precisamos usar a distribuição t de Student em vez da normal, pois a estimativa S_n introduz incerteza adicional. A distribuição t tem caudas mais pesadas que a normal, resultando em intervalos ligeiramente mais largos.

4.3 Exemplo 8: Variância e Desvio Padrão

Exemplo 8: Sejam X_1, \dots, X_n uma amostra de $X \sim N(\mu, \sigma^2)$ com $\mu \in \mathbb{R}$ e $\sigma \in \mathbb{R}_+$ desconhecidos. Encontre o intervalo de confiança $1 - \alpha$ para σ .

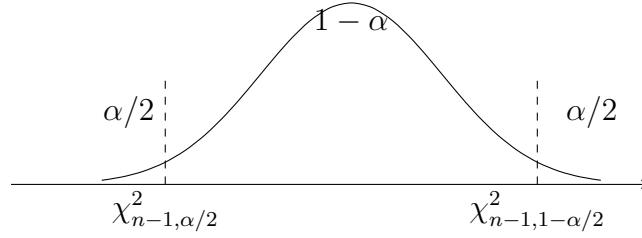
Solução: Note que

$$U = \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}. \quad (43)$$

Justificativa: A estatística $(n-1)S_n^2/\sigma^2 = \sum_{i=1}^n (X_i - \bar{X})^2/\sigma^2$ segue distribuição qui-quadrado com $n-1$ graus de liberdade. Isso decorre do fato de que, embora tenhamos n desvios $(X_i - \bar{X})$, eles satisfazem a restrição $\sum_{i=1}^n (X_i - \bar{X}) = 0$, resultando em $n-1$ graus de liberdade.

Sejam $\chi^2_{n-1, \alpha/2}$ e $\chi^2_{n-1, 1-\alpha/2}$ tais que

$$P(U < \chi^2_{n-1, \alpha/2}) = \frac{\alpha}{2} \quad \text{e} \quad P(U > \chi^2_{n-1, 1-\alpha/2}) = \frac{\alpha}{2}. \quad (44)$$



Então,

$$P(\chi_{n-1,\alpha/2}^2 < U < \chi_{n-1,1-\alpha/2}^2) = 1 - \alpha \quad (45)$$

$$P\left(\chi_{n-1,\alpha/2}^2 < \frac{(n-1)S_n^2}{\sigma^2} < \chi_{n-1,1-\alpha/2}^2\right) = 1 - \alpha \quad (46)$$

Invertendo as desigualdades (lembre-se que ao inverter, as desigualdades se invertem):

$$P\left(\frac{(n-1)S_n^2}{\chi_{n-1,1-\alpha/2}^2} < \sigma^2 < \frac{(n-1)S_n^2}{\chi_{n-1,\alpha/2}^2}\right) = 1 - \alpha \quad (47)$$

Aplicando a raiz quadrada (função crescente, preserva desigualdades):

$$IC_{1-\alpha}(\sigma) = \left[\sqrt{\frac{(n-1)S_n^2}{\chi_{n-1,1-\alpha/2}^2}}, \sqrt{\frac{(n-1)S_n^2}{\chi_{n-1,\alpha/2}^2}} \right]. \quad (48)$$

Observação: Note que a distribuição qui-quadrado não é simétrica, então os quantis $\chi_{n-1,\alpha/2}^2$ e $\chi_{n-1,1-\alpha/2}^2$ não são simétricos em torno da média. Além disso, o intervalo para σ^2 pode ser obtido diretamente sem aplicar a raiz quadrada.

4.4 Tabela Comparativa: Intervalos para População Normal - Uma Amostra

Parâmetro	Condições	Intervalo de Confiança ($1 - \alpha$)
μ	σ^2 conhecido	$\left(\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right)$
μ	σ^2 desconhecido	$\left(\bar{X} \pm t_{n-1,\alpha/2} \frac{S_n}{\sqrt{n}}\right)$
σ^2	μ desconhecido	$\left(\frac{(n-1)S_n^2}{\chi_{n-1,1-\alpha/2}^2}, \frac{(n-1)S_n^2}{\chi_{n-1,\alpha/2}^2}\right)$
σ	μ desconhecido	$\left[\sqrt{\frac{(n-1)S_n^2}{\chi_{n-1,1-\alpha/2}^2}}, \sqrt{\frac{(n-1)S_n^2}{\chi_{n-1,\alpha/2}^2}}\right]$

Tabela 1: Resumo dos intervalos de confiança para uma amostra de população normal

5 Intervalos de Confiança para Duas Amostras

Nesta seção, consideraremos problemas de inferência comparando parâmetros de duas populações normais independentes.

5.1 Abordagem Geral para Duas Amostras

Focaremos na abordagem da quantidade pivotal. Para uma função paramétrica diferenciável $K(\theta)$ para $\theta \in \Theta \subset \mathbb{R}^p$, assuma que temos um estimador $\hat{K}(\hat{\theta})$ que é função de uma estatística suficiente (mínima) para θ .

Frequentemente, a distribuição de

$$U = \frac{\hat{K}(\theta) - K(\theta)}{\hat{\gamma}} \quad (49)$$

não dependerá de θ , $\forall \theta \in \Theta$, para algum $\gamma > 0$.

Se τ é conhecido, podemos obter a, b tais que

$$P_\theta \left(a < \frac{\hat{\kappa}(\theta) - \kappa(\theta)}{\tau} < b \right) = 1 - \alpha \quad (50)$$

Desta última identidade, obtém-se o intervalo de confiança $1 - \alpha$ para $\kappa(\theta)$.

Para τ desconhecido:

$$P_\theta \left(a < \frac{\hat{\kappa}(\theta) - \kappa(\theta)}{\hat{\tau}} < b \right) = 1 - \alpha \quad (51)$$

Desta identidade, obtém-se o intervalo de confiança.

Os resultados anteriores são locação. Quando a inferência é sobre o parâmetro de escala, costuma-se utilizar o pivô

$$U = \frac{\hat{\kappa}(\theta)}{\kappa(\theta)} \quad (52)$$

cuja distribuição geralmente independe de θ . Para este caso, usa-se

$$P_\theta \left(a < \frac{\hat{\kappa}(\theta)}{\kappa(\theta)} < b \right) = 1 - \alpha \quad (53)$$

5.2 Exemplo 9: Diferença de Médias (Variâncias Iguais)

Exemplo 9: Suponha x_{i1}, \dots, x_{in_i} para $i = 1, 2$ duas amostras aleatórias de $X_i \sim N(\mu_i, \sigma^2)$ e independentes entre elas $X_1 \perp X_2$. Vamos assumir que $\theta = (\mu_1, \mu_2, \sigma) \in \mathbb{R}^2 \times \mathbb{R}^+$ é desconhecido, encontrar o intervalo bilateral com confiança $1 - \alpha$ para $u(\theta) = \mu_1 - \mu_2$.

Solução: Pelo teorema da suficiência,

$$T_2 = \left\{ \frac{1}{n_1} \sum_{i=1}^{n_1} x_{1i}, \frac{1}{n_2} \sum_{i=1}^{n_2} x_{2i}, \frac{\sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2 + \sum_{i=1}^{n_2} (x_{2i} - \bar{x}_2)^2}{n_1 + n_2 - 2} \right\} \quad (54)$$

é conjuntamente suficiente para $\theta = (\mu_1, \mu_2, \sigma^2)$.

O termo S_p^2 é chamado de variância amostral conjunta (pooled variance) e pode ser reescrito como:

$$S_1^2 = (n_1 - 1)^{-1} \sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2, \quad S_2^2 = (n_2 - 1)^{-1} \sum_{i=1}^{n_2} (x_{2i} - \bar{x}_2)^2 \quad (55)$$

$$S_p^2 = (n_1 + n_2 - 2)^{-1} [(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2] \quad (56)$$

Justificativa da variância conjunta: Como assumimos que $\sigma_1^2 = \sigma_2^2 = \sigma^2$, combinamos as informações de ambas as amostras para estimar a variância comum. A variância conjunta é uma média ponderada das variâncias amostrais, onde os pesos são os graus de liberdade de cada amostra.

Note que como $(n_1 - 1)S_1^2/\sigma^2 \sim \chi_{n_1-1}^2$ e $(n_2 - 1)S_2^2/\sigma^2 \sim \chi_{n_2-1}^2$ são independentes, então

$$(n_1 + n_2 - 2)S_p^2/\sigma^2 \sim \chi_{n_1+n_2-2}^2 \quad (57)$$

Daí, vale-se:

$$U = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \cdot \sqrt{\frac{n_1 + n_2 - 2}{S_p^2/\sigma^2}} = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2} \quad (58)$$

Justificativa: O numerador $\frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0, 1)$ porque $\bar{x}_1 - \bar{x}_2 \sim N(\mu_1 - \mu_2, \sigma^2(\frac{1}{n_1} + \frac{1}{n_2}))$. O denominador é a raiz quadrada de uma qui-quadrado dividida por seus graus de liberdade, resultando em uma distribuição t .

Para $\nu = n_1 + n_2 - 2$ e $t_{\nu, \alpha/2} > 0$ tal que

$$P(U > t_{\nu, \alpha/2}) = \frac{\alpha}{2}, \quad (59)$$

então

$$P(-t_{\nu, \alpha/2} < U < t_{\nu, \alpha/2}) = 1 - \alpha \Rightarrow \quad (60)$$

$$P\left(-t_{\nu, \alpha/2} < \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < t_{\nu, \alpha/2}\right) = 1 - \alpha \Rightarrow \quad (61)$$

$$P\left(\bar{X}_1 - \bar{X}_2 - t_{\nu, \alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{X}_1 - \bar{X}_2 + t_{\nu, \alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) = 1 - \alpha \quad (62)$$

Dai,

$$IC_{1-\alpha}(\mu_1 - \mu_2) = \left\{ \bar{X}_1 - \bar{X}_2 \pm t_{\nu, \alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\}. \quad (63)$$

5.3 Exemplo 10: Razão de Variâncias

Exemplo 10: Bilateral com confiança $1 - \alpha$ para $\mu(\theta) = \sigma_1^2/\sigma_2^2$

Idéia: Pode-se mostrar (fica como exercício) que

$$\bar{X}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} X_{1i}, \quad \bar{X}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} X_{2i}, \quad S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1)^2, \quad S_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (X_{2i} - \bar{X}_2)^2 \quad (64)$$

são estatísticas suficientes para μ_1, μ_2, σ_1^2 e σ_2^2 respectivamente. Note que (por definição da distribuição F)

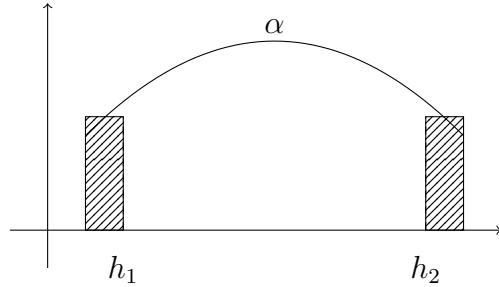
$$U = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1, n_2-1} \quad (65)$$

uma vez que $(n_1 - 1)S_1^2/\sigma_1^2 \sim \chi_{n_1-1}^2$ e $(n_2 - 1)S_2^2/\sigma_2^2 \sim \chi_{n_2-1}^2$ são independentes. Logo, U é uma quantidade pivotal.

Justificativa: A razão entre duas variáveis qui-quadrado independentes, cada uma dividida por seus graus de liberdade, segue distribuição F . Como as variâncias amostrais são independentes (amostras independentes) e cada uma segue uma qui-quadrado quando padronizada, sua razão segue F .

Sejam $h_1 = F_{n_1-1, n_2-1; \alpha/2}$ e $h_2 = F_{n_1-1, n_2-1; 1-\alpha/2}$ quantidades tais que

$$P(U < h_1) = \frac{\alpha}{2} \quad \text{e} \quad P(U > h_2) = \frac{\alpha}{2}.$$



Daí,

$$P_\theta(h_1 < U < h_2) = 1 - \alpha \quad \Rightarrow \quad P_\theta\left(h_1 < \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} < h_2\right) = 1 - \alpha \quad (66)$$

Reescrevendo:

$$P_\theta\left(h_1 < \frac{\sigma_2^2}{\sigma_1^2} \cdot \frac{S_1^2}{S_2^2} < h_2\right) = 1 - \alpha \quad (67)$$

Multiplicando por $\frac{S_2^2}{S_1^2}$ (positivo):

$$P_\theta\left(h_1 \frac{S_2^2}{S_1^2} < \frac{\sigma_2^2}{\sigma_1^2} < h_2 \frac{S_2^2}{S_1^2}\right) = 1 - \alpha \quad (68)$$

Invertendo para obter $\frac{\sigma_1^2}{\sigma_2^2}$:

$$P_\theta\left(\frac{1}{h_2} \frac{S_1^2}{S_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{1}{h_1} \frac{S_1^2}{S_2^2}\right) = 1 - \alpha \quad (69)$$

Usando a propriedade $F_{1-\alpha/2, \nu_1, \nu_2} = \frac{1}{F_{\alpha/2, \nu_2, \nu_1}}$:

$$P_\theta\left(F_{1-\frac{\alpha}{2}, n_2-1, n_1-1}^{-1} \frac{S_1^2}{S_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < F_{1-\frac{\alpha}{2}, n_1-1, n_2-1}^{-1} \frac{S_1^2}{S_2^2}\right) = 1 - \alpha \quad (70)$$

isto é,

$$IC_{1-\alpha}\left(\frac{\sigma_1^2}{\sigma_2^2}\right) = \left(F_{1-\frac{\alpha}{2}, n_2-1, n_1-1}^{-1} \frac{S_1^2}{S_2^2}, F_{1-\frac{\alpha}{2}, n_1-1, n_2-1}^{-1} \frac{S_1^2}{S_2^2}\right) \quad (71)$$

Observação: A distribuição F não é simétrica, e há uma relação importante entre os quantis: $F_{1-\alpha/2, \nu_1, \nu_2} = \frac{1}{F_{\alpha/2, \nu_2, \nu_1}}$. Isso é usado na construção do intervalo.

Parte II

Teste de Razão de Verossimilhanças (TRV)

6 Fundamentos do TRV

Discutimos que pode não existir teste UMP para o caso simples bilateral. O teste da razão entre verossimilhanças proposto por Neyman e Pearson (1928, 1933) é um método útil para lidar com este caso.

6.1 Construção

Sejam x_1, \dots, x_n uma amostra de X com fdp (ou fmpt) dada por $f(x_i; \theta)$, em que $\theta = (\theta_1, \dots, \theta_p)^T \in \Theta \subset \mathbb{R}^p$ é o vetor de parâmetros desconhecidos.

Desejamos testar

$$H_0 : \theta \in \Theta_0 \quad \text{e} \quad H_1 : \theta \in \Theta_1$$

tal que $\Theta = \Theta_0 \cup \Theta_1$ e $\Theta_0 \cap \Theta_1 = \emptyset$ com nível de significância α .

A função de verossimilhança associada é dada por:

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta), \quad \theta \in \Theta \tag{72}$$

Fixamos nossa atenção em

$$\sup_{\theta \in \Theta_0} \{L(\theta)\}$$

interpretada como a melhor evidência em favor de H_0 . Adicionalmente,

$$\sup_{\theta \in \Theta} \{L(\theta)\}$$

representa a melhor evidência em favor de Θ , sem considerar restrição.

A estatística da razão entre verossimilhanças (RV) é dada por:

$$\Lambda = \frac{\sup_{\theta \in \Theta_0} \{L(\theta)\}}{\sup_{\theta \in \Theta} \{L(\theta)\}} \tag{73}$$

A regra de decisão do teste RV (TRV) é dada por:

"Rejeitamos H_0 se, e só se, Λ é pequena ($< k$).

Interpretação: A razão de verossimilhanças compara a máxima verossimilhança sob H_0 com a máxima verossimilhança sem restrições. Se Λ é pequeno, significa que os dados são muito mais prováveis sob H_1 do que sob H_0 , levando à rejeição de H_0 .

Valores pequenos de Λ implicam valores pequenos de $\sup_{\theta \in \Theta_0} L(\theta)$ em comparação com valores de $\sup_{\theta \in \Theta} L(\theta)$.

Note que o $c_{1\alpha}$ e c deve ser definido em $(0, 1)$ tal que o TRV tenha nível α .

6.2 Notas Importantes

(1) No entanto, com frequência se objetiva testar parte dos parâmetros de θ , digamos $\theta_0 = (\theta_1, \dots, \theta_q)^T$ tal que $q < p$, conhecidos como parâmetros de interesse:

$$H_0 : (\theta_1, \dots, \theta_q) = (\theta_{1,0}, \dots, \theta_{q,0})$$

Os demais parâmetros $(\theta_{q+1}, \dots, \theta_p)$ são chamados de parâmetros de perturbação ou incógnitos.

(2) Sobre a derivação de Λ :

$$\sup_{\theta \in \Theta_0} \{L(\theta)\} = L(\hat{\theta}), \quad (74)$$

em que $\hat{\theta}$ representa o estimador de máxima verossimilhança (EMV) restrito (assumindo o parâmetro de interesse conhecido):

$$\hat{\theta} = (\theta_{1,0}, \dots, \theta_{q,0}, \hat{\theta}_{q+1}, \dots, \hat{\theta}_p)$$

Se $(\hat{\theta}_{\text{restr}}, \hat{\theta}_{\text{livre}})$ é o EMV sob H_0 ,

$$\sup_{\theta \in \Theta_0} L(\theta) = L(\hat{\theta}_{\text{restr}}), \quad (75)$$

em que $\hat{\theta}_{\text{restr}}$ é o EMV restrito (considerando apenas Θ_0).

Por outro lado,

$$\sup_{\theta \in \Theta} L(\theta) = L(\hat{\theta}), \quad (76)$$

em que $\hat{\theta}$ é o EMV irrestrito (considerando todo o espaço paramétrico).

7 TRV para Uma Amostra - População Normal

Focaremos nossa atenção sobre a população normal em alguns TRV's. Seguem os casos:

7.1 Caso 1: Média com Variância Conhecida (Teste Z)

Caso 1:

$$\begin{cases} H_0 : \mu = \mu_0 & \text{para } \sigma^2 \text{ conhecido} \\ H_1 : \mu \neq \mu_0 & \end{cases} \quad (77)$$

(Teste Z)

Sejam X_1, \dots, X_n uma a.a. de $X_i \sim N(\mu, \sigma^2)$ com μ desconhecido e σ_0 conhecido. Aqui, temos

$$\Theta = \mathbb{R} \quad \text{e} \quad \Theta_0 = \{\mu_0\}.$$

A função de verossimilhança é

$$L(\mu) = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2 \right\}. \quad (78)$$

Desta última identidade, tem-se

$$\sup_{\mu \in \Theta_0} \{L(\mu)\} = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu_0}{\sigma} \right)^2 \right\} \quad (79)$$

$$\sup_{\mu \in \Theta} \{L(\mu)\} = (2\pi\sigma^2)^{-n/2} \sup_{\mu \in \Theta} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2 \right\} \quad (80)$$

O máximo ocorre quando $\mu = \bar{x}$, pois \bar{x} é o EMV para μ . Portanto:

$$\sup_{\mu \in \Theta} \{L(\mu)\} = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^2 \right\} \quad (81)$$

em que $\bar{x} = \hat{\mu}$ é a estimativa de MV para μ .

Usando a identidade:

$$\sum_{i=1}^n (x_i - c)^2 = \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - c)^2 \quad (82)$$

$$= \sum_{i=1}^n \{(x_i - \bar{x})^2 + 2(x_i - \bar{x})(\bar{x} - c) + (\bar{x} - c)^2\} \quad (83)$$

$$= \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - c)^2 \quad (84)$$

Assim,

$$\Lambda = \exp \left\{ -\frac{1}{2\sigma^2} [n(\bar{x} - \mu_0)^2] \right\} \quad (85)$$

O TRV fica definido por

"Rejeitar H_0 se, e só se, $\Lambda < k'$ "

$$\Leftrightarrow \left\{ n \left(\frac{\bar{x} - \mu_0}{\sigma} \right)^2 > -2 \log k' \right\} \quad (86)$$

$$\Leftrightarrow \left\{ \sqrt{n} \left| \frac{\bar{x} - \mu_0}{\sigma} \right| > \sqrt{-2 \log k'} \right\} \quad (87)$$

$$\Leftrightarrow \{|Z(x_1)| > k''\} \quad (88)$$

em que

$$Z(x_1) = \sqrt{n} \left(\frac{\bar{x} - \mu_0}{\sigma} \right) \quad (89)$$

Note como $k'' = z_{\alpha/2}$ e satisfaz

$$P_{H_0} (|Z| > z_{\alpha/2}) = \alpha \quad (90)$$

Interpretação: O teste rejeita H_0 quando a estatística Z está muito distante de zero, indicando que a média amostral está significativamente diferente de μ_0 .

7.2 Caso 2: Média com Variância Desconhecida (Teste t)

Caso 2:

$$\begin{cases} H_0 : \mu = \mu_0 & \text{para } \sigma^2 \text{ desconhecido} \\ H_1 : \mu \neq \mu_0 \end{cases} \quad (91)$$

(Teste t)

Considere testar $H_0 : \mu = \mu_0$ e $H_1 : \mu \neq \mu_0$ baseado em x_1, \dots, x_n como uma a.a. de $X \sim N(\mu, \sigma^2)$. Neste caso

$$\Theta_0 = \{(\mu_0, \sigma^2); \mu_0 \text{ é fixado e } \sigma \in \mathbb{R}_+\}.$$

Aqui a função de verossimilhança é

$$L(\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \cdot \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2 \right\}. \quad (92)$$

Desta expressão,

$$\sup_{(\mu_0, \sigma^2) \in \Theta_0} L(\mu, \sigma^2) = L(\mu_0, \hat{\sigma}^2) = (2\pi\hat{\sigma}^2)^{-n/2} \cdot \exp \left\{ -\frac{1}{2\hat{\sigma}^2} \sum_{i=1}^n (x_i - \mu_0)^2 \right\}, \quad (93)$$

em que

$$\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n (x_i - \mu_0)^2$$

é a estimativa de MV para σ^2 sob H_0 . Simplificando a última expressão temos:

$$\sup_{(\mu_0, \sigma^2) \in \Theta_0} L(\mu, \sigma^2) = (2\pi\hat{\sigma}^2)^{-n/2} e^{-n/2}. \quad (94)$$

Por outro lado, tem-se:

$$\sup_{(\mu, \sigma^2) \in \Theta} L(\mu, \sigma^2) = L(\bar{x}, \hat{\sigma}^2) = (2\pi\hat{\sigma}^2)^{-n/2} \cdot \exp \left\{ -\frac{1}{2\hat{\sigma}^2} \sum_{i=1}^n (x_i - \bar{x})^2 \right\}. \quad (95)$$

em que $\hat{\mu} = \bar{x}$ e $\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n (x_i - \bar{x})^2$ são os estimadores de MV irrestritos para μ e σ . Simplificando,

$$\sup_{(\mu, \sigma) \in \Theta} L(\mu, \sigma) = (2\pi\hat{\sigma}^2)^{-n/2}. \quad (96)$$

Finalmente,

$$\Lambda = \frac{\sup_{(\mu, \sigma) \in \Theta_0} L(\mu, \sigma)}{\sup_{(\mu, \sigma) \in \Theta} L(\mu, \sigma)} \quad (97)$$

$$= \left\{ \frac{\sum_{i=1}^n (x_i - \mu_0)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right\}^{-n/2} \quad (98)$$

Usando a identidade $\sum_{i=1}^n (x_i - \mu_0)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2$:

$$= \left\{ \frac{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right\}^{-n/2} \quad (99)$$

$$= \left\{ 1 + \frac{n(\bar{x} - \mu_0)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right\}^{-n/2} \quad (100)$$

O TRV fica definido por: dado um nível α , rejeitamos $H_0 \iff \Lambda < k$ ($0 < k < 1$). Logo, como $f(x) = x^{-n/2}$ é estritamente decrescente,

$$R_c = \left\{ 1 + \frac{n(\bar{x} - \mu_0)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} > k^{-2/n} \right\} = \left\{ \frac{n(\bar{x} - \mu_0)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} > k^{-2/n} - 1 \right\} \quad (101)$$

Definindo $s^2 = (n-1)^{-1} \sum_{i=1}^n (x_i - \bar{x})^2$ e manipulando:

$$R_c = \left\{ \frac{\sqrt{n} |\bar{x} - \mu_0|}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} > \sqrt{k'} \right\} \quad (102)$$

Note que:

$$A = \frac{\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}}{\sqrt{\sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^2}} = \frac{\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)s^2}{\sigma^2}}} = \frac{\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}}{\frac{s}{\sigma}} \quad (103)$$

Para $\delta_c = (n-1)^{-1} \sum_{i=1}^n (x_i - \bar{x})^2 = s^2$,

$$R_c = \left\{ (n-1)^{-1/2} |T(x)| > \sqrt{k'} \right\} \quad (104)$$

para $T(x) = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ e $x = (x_1, \dots, x_n)^T$. Logo:

$$R_c = \{|T(x)| > k''\} \quad (105)$$

Note que:

$$T(X) \text{ sob } H_0 \sim t_{n-1}. \quad (106)$$

Seja $t_{n-1,\alpha/2}$ tal que:

$$P_{H_0} (T > t_{n-1,\alpha/2}) = \frac{\alpha}{2}. \quad (107)$$

Então o TRV tem fronteira crítica:

$$\psi(X^1) = \begin{cases} 1, & |T(X^1)| > t_{n-1,\alpha/2} \\ 0, & \text{c.c.} \end{cases}$$

Interpretação: Quando a variância é desconhecida, usamos a distribuição t de Student em vez da normal. A estatística T padroniza a diferença entre a média amostral e μ_0 usando o desvio padrão amostral.

7.3 Caso 3: Variância com Média Conhecida

TRV para σ^2 com μ conhecido

Sejam X_1, \dots, X_n v.a. de $X \sim N(\mu, \sigma^2)$ com média conhecida e $\sigma^2 \in \mathbb{R}^+$ desconhecido. Dado um nível $\alpha \in (0, 1)$, considere a derivação do TRV para

$$H_0 : \sigma^2 = \sigma_0^2 \quad \text{vs} \quad H_1 : \sigma^2 \neq \sigma_0^2.$$

A função de verossimilhança é dada por:

$$l(\sigma^2) = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right\} \quad (108)$$

Daí, temos:

$$\sup_{(\mu, \sigma^2) \in \Theta} l(\sigma^2) = (2\pi\hat{\sigma}^2)^{-n/2} \exp \left\{ -\frac{1}{2\hat{\sigma}^2} \sum_{i=1}^n (x_i - \mu)^2 \right\} \quad (109)$$

em que

$$\Theta_0 = \{(\mu, \sigma^2) \mid \mu \text{ é fixado e } \sigma^2 = \sigma_0^2 \text{ é fixado}\}$$

e $\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n (x_i - \mu)^2$ é o EMV irrestrito.

$$\sup_{(\mu, \sigma^2) \in \Theta_0} l(\sigma^2) = (2\pi\sigma_0^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \mu)^2 \right\} \quad (110)$$

em que

$$\Theta = \{\mu, \sigma\} : \mu \text{ fixado e } \sigma \in \mathbb{R}_+$$

e

$$\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n (x_i - \mu)^2.$$

Finalmente,

$$\begin{aligned} \Lambda &= \left\{ \frac{\hat{\sigma}^2}{\sigma_0^2} \right\}^{n/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma_0^2} + \frac{n}{2} \right\} \\ &= \left[\frac{\hat{\sigma}^2}{\sigma_0^2} \right]^{n/2} \exp \left(1 - \frac{\hat{\sigma}^2}{\sigma_0^2} \right)^{n/2}. \end{aligned} \quad (111)$$

Deste modo, o TRV fica definido como

$$\text{rejeita-se } H_0 \iff \Lambda < k \quad (\alpha \in (0, 1)).$$

Como a função $g(u) = ue^{1-u}$ para $u > 0$ tem máximo em $u = 1$ e é simétrica em torno desse ponto, temos que a região crítica é bilateral. Note que:

$$\lim_{u \rightarrow 0} g(u) = 0 \quad \text{e} \quad \lim_{u \rightarrow \infty} g(u) = 0.$$

A função $g(u)$ atinge seu máximo quando $u = 1$, ou seja, quando $\hat{\sigma}^2 = \sigma_0^2$. Valores muito pequenos ou muito grandes de $\hat{\sigma}^2/\sigma_0^2$ levam a valores pequenos de Λ , indicando evidência contra H_0 .

Logo, para

$$T(X) = \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma_0^2} \sim \chi_n^2, \quad (112)$$

rejeita-se H_0 se, e só se, (para $a, b \in \mathbb{R}$, $a < b$)

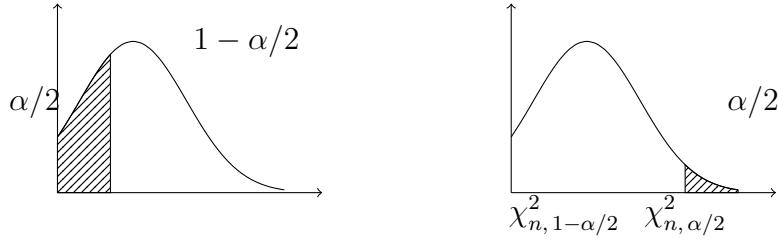
$$[T(X) < a] \quad \text{ou} \quad [T(X) > b]$$

tal que

$$P_{H_0}(T < a) = \frac{\alpha}{2} \quad \Rightarrow \quad a = \chi_{n, 1-\alpha/2}^2 \quad (113)$$

e

$$P_{H_0}(T > b) = \frac{\alpha}{2} \quad \Rightarrow \quad b = \chi_{n, \alpha/2}^2 \quad (114)$$



A função crítica é dada por:

$$\psi(x) = \begin{cases} 1, & T < \chi_{n, 1-\alpha/2}^2 \text{ ou } T > \chi_{n, \alpha/2}^2, \\ 0, & \text{caso contrário.} \end{cases} \quad (115)$$

Interpretação: O teste rejeita H_0 quando a variância amostral é muito pequena ou muito grande em relação a σ_0^2 , indicando evidência de que a variância verdadeira difere de σ_0^2 .

7.4 Caso 4: Variância com Média Desconhecida

Caso 4:

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 & \text{para } \mu \text{ desconhecido,} \\ H_1 : \sigma^2 \neq \sigma_0^2 & \end{cases}$$

(Teste χ_{n-1}^2)

Sejam X_1, \dots, X_n uma a.a. de $X_i \sim N(\mu, \sigma^2)$ com μ desconhecido e σ_0 conhecido. Aqui, temos

$$\Theta = \mathbb{R} \times \mathbb{R}_+ \quad \text{e} \quad \Theta_0 = \mathbb{R} \times \{\sigma_0^2\}.$$

A função de verossimilhança é

$$L(\mu) = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2 \right\}. \quad (116)$$

Desta última identidade, tem-se

$$\sup_{\mu \in \Theta_0} \{L(\mu)\} = (2\pi\sigma^2)^{-n/2} \sup_{\mu \in \Theta_0} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2 \right\} \quad (117)$$

$$\sup_{\mu \in \Theta} \{L(\mu)\} = (2\pi\sigma^2)^{-n/2} \sup_{\mu \in \Theta} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^2 \right\} \quad (118)$$

em que $\bar{x} = \hat{\mu}$ é a estimativa de MV para μ .

A derivação segue de forma similar ao Caso 3, mas agora usando $(n-1)S_n^2/\sigma_0^2 \sim \chi_{n-1}^2$ em vez de χ_n^2 , pois a média é estimada a partir dos dados.

Observação: A diferença crucial entre os Casos 3 e 4 é o número de graus de liberdade da distribuição qui-quadrado. Quando a média é conhecida, usamos χ_n^2 ; quando é desconhecida e estimada, perdemos um grau de liberdade, resultando em χ_{n-1}^2 .

8 TRV para Duas Amostras

Nesta seção, vamos discutir problemas relacionados ao TRV para duas amostras em dois contextos principais:

$$\text{para } X \sim N(\mu_x, \sigma_x^2) \quad \text{e} \quad Y \sim N(\mu_y, \sigma_y^2)$$

8.1 TRV para Comparar Médias

Sejam x_1, \dots, x_n e y_1, \dots, y_m a.a.s independentes de $X \sim N(\mu_x, \sigma_x^2)$ e $Y \sim N(\mu_y, \sigma_y^2)$, respectivamente.

Assuma que $\sigma_x^2 = \sigma_y^2 = \sigma^2 \in \mathbb{R}_+$ e $\theta = (\mu_x, \mu_y, \sigma^2) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+$ é desconhecido.

Dado $\alpha \in (0, 1)$ como o nível, vamos derivar o TRV para

$$H_0 : \mu_x = \mu_y \quad \text{vs} \quad H_1 : \mu_x \neq \mu_y \quad (119)$$

Inicialmente, lembramos que:

$$\begin{aligned} \bar{X} &= n^{-1} \sum_{i=1}^n x_i, & \bar{Y} &= m^{-1} \sum_{i=1}^m y_i, & S_X^2 &= (n-1)^{-1} \sum_{i=1}^n (x_i - \bar{X})^2, \\ S_Y^2 &= (m-1)^{-1} \sum_{i=1}^m (y_i - \bar{Y})^2, \\ S^2 &= \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2} \end{aligned}$$

São estimadores para μ_X , μ_Y , σ_X^2 , σ_Y^2 e σ^2 , respectivamente. Neste caso,

$$\begin{aligned} \Theta_0 &= \{(\mu_X, \mu_Y, \sigma^2) : \mu_X = \mu_Y = \mu \in \mathbb{R} \text{ e } \sigma^2 \in \mathbb{R}_+\} \\ \Theta &= \{(\mu_X, \mu_Y, \sigma^2) : \mu_X, \mu_Y \in \mathbb{R} \text{ e } \sigma^2 > 0\} \end{aligned}$$

A verossimilhança é dada por:

$$L(\mu_X, \mu_Y, \sigma^2) \triangleq L(\mu_X, \mu_Y, \sigma^2; x^n, y^m) = (2\pi\sigma^2)^{-\frac{n+m}{2}} \cdot \exp \left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^n (x_i - \mu_X)^2 + \sum_{j=1}^m (y_j - \mu_Y)^2 \right] \right\} \quad (120)$$

para $(\mu_X, \mu_Y, \sigma^2) \in \mathbb{R}^2 \times \mathbb{R}_+$, em que x^n é uma amostra de X e y^m é uma amostra de Y . Assim,

$$\sup_{\{L(\mu_X, \mu_Y, \sigma^2)\}} = \sup \left\{ (2\pi\sigma^2)^{-\frac{n+m}{2}} \cdot \exp \left[-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (x_i - \mu)^2 + \sum_{j=1}^m (y_j - \mu)^2 \right) \right] \right\} \quad (121)$$

$$\text{para } (\mu_X, \mu_Y, \sigma^2) \in \Theta_0$$

De MV restrito para μ e σ^2 são dados por:

$$\hat{\mu} = \frac{n\bar{x} + m\bar{y}}{n+m}, \quad \hat{\sigma}^2 = \frac{(n-1)\Delta_{x;\mu}^2 + (m-1)\Delta_{y;\mu}^2}{n+m} \quad (122)$$

em que

$$\Delta_{x;\mu}^2 = (n-1)^{-1} \sum_{i=1}^n (x_i - \mu)^2, \quad \Delta_{y;\mu}^2 = (m-1)^{-1} \sum_{i=1}^m (y_i - \mu)^2 \quad (123)$$

Aplicando (2) em (1), temos que:

$$\sup\{L(\mu_x, \mu_y, \sigma^2)\} = (2\pi\hat{\sigma}_c^2)^{-\frac{mn}{2}} \cdot e^{-\frac{mn}{2}}, \quad (\mu_x, \mu_y, \sigma^2) \in \Theta_0 \quad (124)$$

Por outro lado, temos que:

$$\begin{aligned} \sup\{L(\mu_x, \mu_y, \sigma^2)\} &= (2\pi\hat{\sigma}^2)^{-\frac{n+m}{2}} \cdot \exp \left\{ -\frac{1}{2\hat{\sigma}^2} \sum_{i=1}^n (x_i - \hat{\mu}_x)^2 + \sum_{i=1}^m (y_i - \hat{\mu}_y)^2 \right\} \\ &= (2\pi\hat{\sigma}^2)^{-\frac{n+m}{2}} \cdot e^{-\frac{n+m}{2}}, \end{aligned} \quad (125)$$

uma vez que $\hat{\mu}_x = \bar{x}$, $\hat{\mu}_y = \bar{y}$ e

$$\hat{\sigma}_{x,y}^2 = \frac{(n-1)\Delta_x^2 + (m-1)\Delta_y^2}{n+m}, \quad (126)$$

para s_x^2 como uma observação de S_x^2 e s_y^2 como uma observação de S_y^2 , são estimativas de MV irrestritas para μ_x , μ_y e σ^2 .

Finalmente,

$$\Lambda = \left\{ \frac{\hat{\sigma}_c^2}{\hat{\sigma}^2} \right\}^{\frac{n+m}{2}} = \left\{ \frac{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^m (y_i - \bar{y})^2}{\frac{n+m}{n+m} \sum_{i=1}^n (x_i - \hat{\mu})^2 + \sum_{i=1}^m (y_i - \hat{\mu})^2} \right\}^{\frac{n+m}{2}}. \quad (127)$$

Note que

$$\begin{aligned} \sum_{i=1}^n (x_i - \hat{\mu}_x)^2 + \sum_{i=1}^m (y_i - \hat{\mu}_y)^2 &= \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \hat{\mu})^2 + \sum_{i=1}^m (y_i - \bar{y})^2 + m(\bar{y} - \hat{\mu})^2 \quad (128) \\ &= \underbrace{(n-1)s_x^2}_{\text{variância amostral de } x} + \underbrace{(m-1)s_y^2}_{\text{variância amostral de } y} + n(\bar{x} - \hat{\mu})^2 + m(\bar{y} - \hat{\mu})^2 \end{aligned}$$

E, para

$$R(\bar{x}, \bar{y}) = \frac{n(\bar{x} - \hat{\mu})^2 + m(\bar{y} - \hat{\mu})^2}{(n-1)s_x^2 + (m-1)s_y^2}, \quad (129)$$

temos

$$\Lambda = \left\{ \frac{1}{1 + R(\bar{x}, \bar{y})} \right\}^{\frac{n+m}{2}} = \{1 + R(\bar{x}, \bar{y})\}^{-\frac{n+m}{2}}. \quad (130)$$

Ainda,

$$n(\bar{x} - \hat{\mu})^2 + m(\bar{y} - \hat{\mu})^2 = n \left[\bar{x} - \frac{n\bar{x} + m\bar{y}}{n+m} \right]^2 + m \left[\bar{y} - \frac{n\bar{x} + m\bar{y}}{n+m} \right]^2 \quad (131)$$

$$= n \left[\frac{m\bar{x} - m\bar{y}}{n+m} \right]^2 + m \left[\frac{n\bar{y} - n\bar{x}}{n+m} \right]^2 \quad (132)$$

$$= \frac{nm^2}{(n+m)^2} (\bar{x} - \bar{y})^2 + \frac{mn^2}{(n+m)^2} (\bar{x} - \bar{y})^2 = \frac{nm}{n+m} (\bar{x} - \bar{y})^2. \quad (133)$$

Então

$$R(\bar{x}, \bar{y}) = \frac{\frac{nm}{n+m}(\bar{x} - \bar{y})^2}{(n-1)s_x^2 + (m-1)s_y^2} = \frac{nm}{n+m} \cdot \frac{(\bar{x} - \bar{y})^2}{(n-1)s_x^2 + (m-1)s_y^2}. \quad (134)$$

A região crítica do TRV é dada por: para

$$z = (\bar{x}^T, \bar{y}^T)^T \in \mathbb{R}^n \times \mathbb{R}^m, \text{ o qual}$$

$$R_c = \{z \in \mathcal{X}; \Lambda(z) < k\} = \left\{ [1 + R(\bar{x}, \bar{y})]^{-\frac{nm}{2}} < k \right\} \Rightarrow \quad (135)$$

$$R_c = \left\{ [1 + R(\bar{x}, \bar{y})] > k^{-\frac{2}{nm}} \right\} \Rightarrow \quad (136)$$

$$R_c = \left\{ R(\bar{x}, \bar{y}) > k^{-\frac{2}{nm}} - 1 \right\} \Rightarrow \quad (137)$$

$$R_c = \left\{ \sqrt{\frac{nm}{n+m} \cdot \frac{(\bar{x} - \bar{y})^2}{(n-1)s_x^2 + (m-1)s_y^2}} > \sqrt{k^{-\frac{2}{nm}} - 1} \right\} \Rightarrow \quad (138)$$

$$R_c = \{|T(\bar{x}, \bar{y})| > k'\}.$$

Note que, para as a.o.s \bar{x} e \bar{y} ,

$$T(\bar{x}, \bar{y}) = \sqrt{\frac{nm}{n+m}} \cdot \frac{\bar{x} - \bar{y}}{\sqrt{(n-1)s_x^2 + (m-1)s_y^2}} = \frac{\bar{x} - \bar{y}}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}} \quad (139)$$

em que $S_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}$ é a variância amostral conjunta.
Ainda, a R_c do TRV pode ser dada por:

$$R_c = \{z \in \mathcal{X} \mid |T(\bar{x}, \bar{y})| > k''\} \quad (140)$$

Podemos ver que

$$T(\bar{x}, \bar{y}) \stackrel{H_0}{\sim} t_{nm-2} \quad (141)$$

Tomando $k'' = t_{nm-2, \frac{\alpha}{2}}$ tal que

$$P_{H_0}(|T(\bar{x}, \bar{y})| > t_{nm-2, \frac{\alpha}{2}}) = \alpha, \quad (142)$$

temos que a função crítica do TRV é dada por:

$$\psi(\bar{x}, \bar{y}) = \begin{cases} 1, & |T(\bar{x}, \bar{y})| > t_{nm-2, \frac{\alpha}{2}}, \\ 0, & \text{caso contrário.} \end{cases} \quad (143)$$

em que

$$T(\bar{x}, \bar{y}) = \frac{\bar{x} - \bar{y}}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}} \quad (144)$$

e

$$S_p = \sqrt{\frac{(n-1)s_x^2 + (m-1)s_y^2}{nm-2}} \quad (145)$$

é uma observação da variância amostral conjunta.

Interpretação: Este teste compara as médias de duas populações normais independentes com variâncias iguais. A estatística T segue distribuição t de Student com $n+m-2$ graus de liberdade sob H_0 , pois combinamos as informações de ambas as amostras para estimar a variância comum.

8.2 TRV para Comparar Variâncias

8.2.1 TRV para $H_0 : \sigma_x^2 = \sigma_y^2$

Sejam x_1, \dots, x_n e y_1, \dots, y_m a.a.s. de $X \sim N(\mu_x, \sigma_x^2)$ e $Y \sim N(\mu_y, \sigma_y^2)$. Assuma que

$$\theta = (\mu_x, \mu_y, \sigma_x^2, \sigma_y^2)^T \in \mathbb{R}^2 \times \mathbb{R}_+^2$$

é desconhecido. Dado que $\alpha \in (0, 1)$ como nível de significância, vamos derivar o TRV para

$$H_0 : \sigma_x^2 = \sigma_y^2 \quad \text{e} \quad H_1 : \sigma_x^2 \neq \sigma_y^2.$$

A verossimilhança é dada por:

$$\begin{aligned} L(\theta) &= L(\theta; \vec{x}, \vec{y}) \\ &= (2\pi)^{-\frac{n+m}{2}} \cdot (\sigma_x^n \sigma_y^m)^{-1} \cdot \exp \left\{ -\frac{1}{2} \left[\sum_{i=1}^n \left(\frac{x_i - \mu_x}{\sigma_x} \right)^2 + \sum_{i=1}^m \left(\frac{y_i - \mu_y}{\sigma_y} \right)^2 \right] \right\}. \end{aligned} \quad (146)$$

Em que \vec{x} é uma a.o. de X e \vec{y} uma a.o. de Y . Note que o estimador de máxima verossimilhança restrito para μ_x, μ_y e σ é dado por:

$$\hat{\mu}_x = \bar{X}, \quad \hat{\mu}_y = \bar{Y}, \quad \hat{\sigma}^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m}$$

tal que

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2, \quad S_Y^2 = \frac{1}{m-1} \sum_{i=1}^m (y_i - \bar{Y})^2$$

respectivamente.

Usualmente, os estimadores invariantes de MV para $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2$ são dados por:

$$\hat{\mu}_x = \bar{X}, \quad \hat{\mu}_y = \bar{Y}, \quad \hat{\sigma}_x^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2, \quad \hat{\sigma}_y^2 = \frac{1}{m} \sum_{i=1}^m (Y_i - \bar{Y})^2 \quad (147)$$

respectivamente. Aqui,

$$\Theta_0 = \{(\mu_x, \mu_y, \sigma_x, \sigma_y) : \mu_x, \mu_y \in \mathbb{R} \text{ e } \sigma_x = \sigma_y \in \mathbb{R}_+\} \quad (148)$$

$$\Theta = \{(\mu_x, \mu_y, \sigma_x, \sigma_y) : \mu_x, \mu_y \in \mathbb{R} \text{ e } \sigma_x^2, \sigma_y^2 \in \mathbb{R}_+\} \quad (149)$$

Dado,

$$\sup_{\Theta_0} \{L(\Theta)\} = (2\pi)^{-\frac{n+m}{2}} (\tilde{\sigma}^2)^{-\frac{n+m}{2}} \cdot \exp \left\{ -\frac{1}{2\tilde{\sigma}^2} [(n-1)s_x^2 + (m-1)s_y^2] \right\} \quad (150)$$

em que s_x^2 é uma observação de S_x^2 e s_y^2 é uma observação de S_y^2 . Logo,

$$\sup_{\Theta_0} \{L(\Theta)\} = (2\pi \tilde{\sigma}^2)^{-\frac{n+m}{2}} e^{-\frac{n+m}{2}} \quad (151)$$

$$\sup_{\Theta} \{L(\Theta)\} = (2\pi)^{-\frac{n+m}{2}} (\hat{\sigma}_x^n \hat{\sigma}_y^m)^{-\frac{1}{2}} e^{-\frac{n+m}{2}} \quad (152)$$

Dado, temos que

$$\Lambda = \frac{(\hat{\sigma}_x^2)^{\frac{n}{2}} (\hat{\sigma}_y^2)^{\frac{m}{2}}}{(\tilde{\sigma}^2)^{\frac{n+m}{2}}} = \frac{(\hat{\sigma}_x^2)^{\frac{n}{2}} (\hat{\sigma}_y^2)^{\frac{m}{2}}}{[(n+m)^{-1}(n\hat{\sigma}_x^2 + m\hat{\sigma}_y^2)]^{\frac{n+m}{2}}} \quad (153)$$

Definindo $\lambda = \frac{n}{n+m}$, tem-se:

$$\Lambda = \left[\frac{\hat{\sigma}_x^2}{\lambda\hat{\sigma}_x^2 + (1-\lambda)\hat{\sigma}_y^2} \right]^{\frac{n}{2}} \left[\frac{\hat{\sigma}_y^2}{\lambda\hat{\sigma}_x^2 + (1-\lambda)\hat{\sigma}_y^2} \right]^{\frac{m}{2}} \quad (154)$$

$$= \hat{\sigma}_y^2 \left[\frac{\hat{\sigma}_x^2/\hat{\sigma}_y^2}{\lambda(\hat{\sigma}_x^2/\hat{\sigma}_y^2) + (1-\lambda)} \right]^{\frac{n}{2}} \left[\frac{1}{\lambda(\hat{\sigma}_x^2/\hat{\sigma}_y^2) + (1-\lambda)} \right]^{\frac{m}{2}} \quad (155)$$

$$\Lambda = \lambda^{-\frac{n+m}{2}} \left(\frac{\hat{\sigma}_x^2/\hat{\sigma}_y^2}{\hat{\sigma}_x^2/\hat{\sigma}_y^2 + b} \right)^{\frac{n+m}{2}}, \quad b \triangleq \frac{1-\lambda}{\lambda} = \frac{m}{n} \quad (156)$$

A RC de TRV é dada por: para $z = (\hat{\sigma}_x^2, \hat{\sigma}_y^2) \in \mathbb{R}^n \times \mathbb{R}^m$

$$\{z \in \mathbb{R}^n \times \mathbb{R}^m, \Lambda(z) < c'\}, \quad u \text{ escalar} \quad (157)$$

$$\left\{ \frac{(\hat{\sigma}_x^2/\hat{\sigma}_y^2)^{n/2}}{[(\hat{\sigma}_x^2/\hat{\sigma}_y^2) + b]^{\frac{n+m}{2}}} < \frac{k \cdot \lambda^{-\frac{n+m}{2}}}{c'} \right\} \quad (158)$$

A fim de representar a RV de modo implementável, vamos proceder como segue: considere a função

$$g(u) = u^{n/2}(u+b)^{-\frac{n+m}{2}}, \quad u > 0 \quad (159)$$

Vamos investigar o comportamento de Λ . Note que:

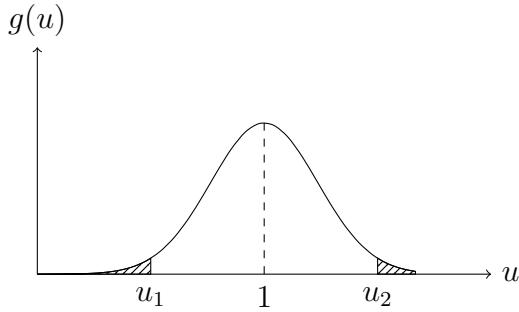
$$\begin{aligned} g'(u) &= \frac{n}{2}u^{\frac{n}{2}-1}(u+b)^{-\frac{n+m}{2}} - u^{\frac{n}{2}}\frac{n+m}{2}(u+b)^{-\frac{n+m}{2}-1} \\ &= u^{\frac{n}{2}-1}(u+b)^{-\frac{n+m}{2}-1} \left[\frac{n}{2}(u+b) - u\frac{n+m}{2} \right] \\ &= u^{\frac{n}{2}-1}(u+b)^{-\frac{n+m}{2}-1} \left[\frac{nb-mu}{2} \right] \end{aligned} \quad (160)$$

Daí, $g(u) \geq 0$ e

$$g'(u) = 0 \Rightarrow u = \frac{nb}{m} = \frac{n \cdot m/n}{m} = 1.$$

Além disso, $\lim_{u \rightarrow 0} g(u) = 0$ e $\lim_{u \rightarrow \infty} g(u) = 0$.

Logo, $g(u)$ é muito pequena quando $u = \frac{\hat{\sigma}_x^2}{\hat{\sigma}_y^2}$ é muito pequena ou muito grande, indicando evidência contra H_0 .



Tomar: A estatística de teste pode ser baseada na razão das variâncias amostrais. Usando a propriedade de que $\frac{S_x^2/\sigma_x^2}{S_y^2/\sigma_y^2} \sim F_{n-1,m-1}$ quando as amostras são independentes, temos que sob H_0 :

$$F = \frac{S_x^2}{S_y^2} \sim F_{n-1,m-1} \quad (161)$$

A função crítica é dada por:

$$\psi(x, y) = \begin{cases} 1, & F < F_{n-1,m-1,1-\alpha/2} \text{ ou } F > F_{n-1,m-1,\alpha/2}, \\ 0, & \text{caso contrário.} \end{cases} \quad (162)$$

Interpretação: Este teste compara as variâncias de duas populações normais independentes. A estatística F segue distribuição F de Fisher com $(n - 1, m - 1)$ graus de liberdade sob H_0 . Rejeitamos H_0 quando a razão das variâncias amostrais é muito pequena ou muito grande.

9 Testes Assintóticos: Wald e Escore

Além do TRV, existem outros métodos assintóticos para construção de testes de hipóteses que são particularmente úteis quando a distribuição exata é difícil de obter ou quando trabalhamos com grandes amostras. Os testes de Wald e Escore são baseados em propriedades assintóticas dos estimadores de máxima verossimilhança.

9.1 Preliminares: Propriedades Assintóticas do EMV

Antes de apresentar os testes de Wald e Escore, recordemos algumas propriedades importantes do estimador de máxima verossimilhança (EMV).

Sejam X_1, \dots, X_n uma amostra aleatória de uma distribuição com função de densidade (ou probabilidade) $f(x; \theta)$, onde $\theta \in \Theta \subset \mathbb{R}^p$ é o vetor de parâmetros. Sob condições de regularidade, o EMV $\hat{\theta}_n$ satisfaz:

1. **Consistência:** $\hat{\theta}_n \xrightarrow{p} \theta_0$ quando $n \rightarrow \infty$, onde θ_0 é o verdadeiro valor do parâmetro.
2. **Normalidade Assintótica:**

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I(\theta_0)^{-1}) \quad (163)$$

onde $I(\theta)$ é a matriz de informação de Fisher, definida por:

$$I(\theta) = E_\theta \left[\frac{\partial \log f(X; \theta)}{\partial \theta} \frac{\partial \log f(X; \theta)}{\partial \theta^T} \right] = -E_\theta \left[\frac{\partial^2 \log f(X; \theta)}{\partial \theta \partial \theta^T} \right] \quad (164)$$

3. Eficiência Assintótica: O EMV é assintoticamente eficiente, atingindo o limite inferior de Cramér-Rao.

Interpretação: A matriz de informação de Fisher $I(\theta)$ mede a quantidade de informação sobre o parâmetro θ contida na amostra. Quanto maior a informação, menor a variância assintótica do estimador.

9.2 Teste de Wald

O teste de Wald é baseado na distribuição assintótica do EMV. A ideia central é que, sob H_0 , o EMV deve estar próximo do valor hipotético, e a distância padronizada segue uma distribuição normal assintótica.

9.2.1 Construção do Teste de Wald

Considere o problema de testar

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

onde $\theta \in \mathbb{R}^p$.

A estatística de Wald é definida como:

$$W_n = n(\hat{\theta}_n - \theta_0)^T I(\hat{\theta}_n)(\hat{\theta}_n - \theta_0) \quad (165)$$

onde $I(\hat{\theta}_n)$ é a matriz de informação de Fisher avaliada no EMV.

Observação: Na prática, frequentemente usamos a matriz de informação observada:

$$J_n(\hat{\theta}_n) = -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \log f(X_i; \hat{\theta}_n)}{\partial \theta \partial \theta^T} \quad (166)$$

em vez de $I(\hat{\theta}_n)$, pois é mais fácil de calcular.

9.2.2 Distribuição Assintótica da Estatística de Wald

Teorema 1: Sob $H_0 : \theta = \theta_0$ e condições de regularidade, temos que:

$$W_n \xrightarrow{d} \chi_p^2 \quad (167)$$

quando $n \rightarrow \infty$, onde p é a dimensão do vetor θ .

Demonstração: Pela normalidade assintótica do EMV, temos:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} Z \sim N(0, I(\theta_0)^{-1}) \quad (168)$$

Pela consistência do EMV, $I(\hat{\theta}_n) \xrightarrow{p} I(\theta_0)$. Usando o teorema de Slutsky:

$$\sqrt{n}I(\hat{\theta}_n)^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{d} I(\theta_0)^{1/2}Z \sim N(0, I_p) \quad (169)$$

onde I_p é a matriz identidade $p \times p$ e $I(\theta_0)^{1/2}$ é a raiz quadrada matricial de $I(\theta_0)$.

Portanto,

$$W_n = n(\hat{\theta}_n - \theta_0)^T I(\hat{\theta}_n)(\hat{\theta}_n - \theta_0) = \left\| \sqrt{n}I(\hat{\theta}_n)^{1/2}(\hat{\theta}_n - \theta_0) \right\|^2 \xrightarrow{d} \|Z'\|^2 \sim \chi_p^2 \quad (170)$$

onde $Z' \sim N(0, I_p)$ e $\|Z'\|^2$ segue distribuição qui-quadrado com p graus de liberdade.

Interpretação: A estatística de Wald mede a distância quadrática entre o EMV e o valor hipotético, ponderada pela informação de Fisher. Valores grandes de W_n indicam que o EMV está longe de θ_0 , fornecendo evidência contra H_0 .

9.2.3 Regra de Decisão

A função crítica do teste de Wald é dada por:

$$\psi_W(x) = \begin{cases} 1, & W_n > \chi_{p,\alpha}^2, \\ 0, & \text{caso contrário.} \end{cases} \quad (171)$$

onde $\chi_{p,\alpha}^2$ é o quantil $(1 - \alpha)$ da distribuição qui-quadrado com p graus de liberdade.

9.2.4 Exemplo: Teste de Wald para Média de População Normal

Sejam X_1, \dots, X_n uma amostra de $X \sim N(\mu, \sigma^2)$ com σ^2 conhecido. Queremos testar $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$.

O EMV para μ é $\hat{\mu}_n = \bar{X}_n$. A informação de Fisher é:

$$I(\mu) = \frac{n}{\sigma^2} \quad (172)$$

A estatística de Wald é:

$$W_n = n(\hat{\mu}_n - \mu_0)^2 \cdot \frac{1}{\sigma^2} = \frac{n(\bar{X}_n - \mu_0)^2}{\sigma^2} \quad (173)$$

Note que $W_n = Z^2$, onde $Z = \sqrt{n}(\bar{X}_n - \mu_0)/\sigma \sim N(0, 1)$ sob H_0 . Portanto, $W_n \sim \chi_1^2$ (exatamente, não apenas assintoticamente).

Observação: Neste caso particular, a distribuição é exata, não apenas assintótica, pois a distribuição normal tem propriedades especiais.

9.3 Teste de Escore (Rao)

O teste de Escore, também conhecido como teste de Rao, é baseado na função escore (score function), que é o gradiente do logaritmo da função de verossimilhança.

9.3.1 Função Escore

A função escore é definida como:

$$U_n(\theta) = \frac{\partial \ell_n(\theta)}{\partial \theta} = \sum_{i=1}^n \frac{\partial \log f(X_i; \theta)}{\partial \theta} \quad (174)$$

onde $\ell_n(\theta) = \sum_{i=1}^n \log f(X_i; \theta)$ é o logaritmo da função de verossimilhança.

Propriedades importantes:

1. $E_\theta[U_n(\theta)] = 0$ (propriedade fundamental da função escore)
2. $\text{Var}_\theta[U_n(\theta)] = nI(\theta)$

9.3.2 Construção do Teste de Escore

Para testar $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$, a estatística de Escore é definida como:

$$S_n = \frac{1}{n} U_n(\theta_0)^T I(\theta_0)^{-1} U_n(\theta_0) \quad (175)$$

ou, equivalentemente,

$$S_n = n^{-1} U_n(\theta_0)^T I(\theta_0)^{-1} U_n(\theta_0) \quad (176)$$

9.3.3 Distribuição Assintótica da Estatística de Escore

Teorema 2: Sob $H_0 : \theta = \theta_0$ e condições de regularidade, temos que:

$$nS_n = U_n(\theta_0)^T I(\theta_0)^{-1} U_n(\theta_0) \xrightarrow{d} \chi_p^2 \quad (177)$$

quando $n \rightarrow \infty$.

Demonstração: Pelo Teorema Central do Limite aplicado à função escore, temos:

$$\frac{1}{\sqrt{n}} U_n(\theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial \log f(X_i; \theta_0)}{\partial \theta} \xrightarrow{d} N(0, I(\theta_0)) \quad (178)$$

Isso decorre do fato de que:

$$E_{\theta_0} \left[\frac{\partial \log f(X_i; \theta_0)}{\partial \theta} \right] = 0 \quad (179)$$

$$\text{Var}_{\theta_0} \left[\frac{\partial \log f(X_i; \theta_0)}{\partial \theta} \right] = I(\theta_0) \quad (180)$$

Portanto,

$$\frac{1}{\sqrt{n}} I(\theta_0)^{-1/2} U_n(\theta_0) \xrightarrow{d} Z \sim N(0, I_p) \quad (181)$$

e consequentemente,

$$nS_n = U_n(\theta_0)^T I(\theta_0)^{-1} U_n(\theta_0) = \|I(\theta_0)^{-1/2} U_n(\theta_0)\|^2 \xrightarrow{d} \|Z\|^2 \sim \chi_p^2 \quad (182)$$

Interpretação: A estatística de Escore mede o tamanho do vetor escore padronizado. Se H_0 é verdadeira, esperamos que a função escore seja próxima de zero (pois $E[U_n(\theta_0)] = 0$). Valores grandes de S_n indicam que a função escore está longe de zero, fornecendo evidência contra H_0 .

9.3.4 Regra de Decisão

A função crítica do teste de Escore é dada por:

$$\psi_S(x) = \begin{cases} 1, & nS_n > \chi_{p,\alpha}^2, \\ 0, & \text{caso contrário.} \end{cases} \quad (183)$$

9.3.5 Exemplo: Teste de Escore para Média de População Normal

Para $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ com σ^2 conhecido, testando $H_0 : \mu = \mu_0$:

A função escore é:

$$U_n(\mu) = \frac{\partial}{\partial \mu} \sum_{i=1}^n \log \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(X_i - \mu)^2}{2\sigma^2} \right) \right] = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu) \quad (184)$$

Avaliando em μ_0 :

$$U_n(\mu_0) = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu_0) = \frac{n}{\sigma^2} (\bar{X}_n - \mu_0) \quad (185)$$

Como $I(\mu) = n/\sigma^2$, temos:

$$nS_n = U_n(\mu_0)^2 \cdot \frac{\sigma^2}{n} = \frac{n^2 (\bar{X}_n - \mu_0)^2}{\sigma^4} \cdot \frac{\sigma^2}{n} = \frac{n (\bar{X}_n - \mu_0)^2}{\sigma^2} \quad (186)$$

Portanto, $nS_n = W_n$ neste caso, mostrando que os testes de Wald e Escore coincidem para este exemplo.

9.4 Relação entre TRV, Wald e Escore

Os três testes (TRV, Wald e Escore) são assintoticamente equivalentes sob condições de regularidade. Especificamente, temos:

Teorema 3: Sob H_0 e condições de regularidade, as estatísticas $-2 \log \Lambda_n$, W_n e nS_n são assintoticamente equivalentes, no sentido de que:

$$-2 \log \Lambda_n = W_n + o_p(1) = nS_n + o_p(1) \quad (187)$$

e todas convergem em distribuição para χ_p^2 .

Interpretação: Para grandes amostras, os três testes fornecerão resultados similares. A escolha entre eles pode depender de:

- **TRV:** Requer estimativa sob H_0 e H_1 , mas é invariante a reparametrizações.
- **Wald:** Requer apenas o EMV irrestrito, mas não é invariante a reparametrizações.
- **Escore:** Requer apenas avaliação da função escore em θ_0 , mas também não é invariante a reparametrizações.

9.5 Teste de Wald para Hipóteses Parciais

Frequentemente, queremos testar apenas parte dos parâmetros. Considere $\theta = (\theta_1, \theta_2)^T$ e o teste:

$$H_0 : \theta_1 = \theta_{1,0} \quad \text{vs} \quad H_1 : \theta_1 \neq \theta_{1,0}$$

onde θ_2 é um parâmetro de perturbação.

Seja $\hat{\theta}_n = (\hat{\theta}_{1,n}, \hat{\theta}_{2,n})^T$ o EMV irrestrito. A estatística de Wald é:

$$W_n = n(\hat{\theta}_{1,n} - \theta_{1,0})^T [I^{11}(\hat{\theta}_n)]^{-1} (\hat{\theta}_{1,n} - \theta_{1,0}) \quad (188)$$

onde $I^{11}(\theta)$ é o bloco $(1, 1)$ da inversa da matriz de informação de Fisher, correspondente a θ_1 .

Sob H_0 , temos $W_n \xrightarrow{d} \chi_{p_1}^2$, onde p_1 é a dimensão de θ_1 .

10 Teste de Wilcoxon

O teste de Wilcoxon é um teste não-paramétrico (livre de distribuição) que não requer suposições sobre a forma da distribuição subjacente. Existem duas versões principais: o teste de postos sinalizados de Wilcoxon (uma amostra) e o teste de soma de postos de Wilcoxon (duas amostras).

10.1 Teste de Wilcoxon para Uma Amostra (Postos Sinalizados)

Este teste é usado para testar se a mediana de uma população é igual a um valor especificado, sem assumir normalidade.

10.1.1 Hipóteses e Suposições

Sejam X_1, \dots, X_n uma amostra aleatória de uma distribuição contínua e simétrica em torno da mediana M . Queremos testar:

$$H_0 : M = M_0 \quad \text{vs} \quad H_1 : M \neq M_0$$

Suposição de simetria: A distribuição é simétrica em torno de M , ou seja, $X - M$ e $-(X - M)$ têm a mesma distribuição.

10.1.2 Construção da Estatística

O procedimento é o seguinte:

1. Calcule as diferenças: $D_i = X_i - M_0$, $i = 1, \dots, n$.
2. Remova observações com $D_i = 0$ (se houver). Seja n' o número de observações restantes.
3. Atribua postos (ranks) R_i aos valores absolutos $|D_i|$: o menor $|D_i|$ recebe posto 1, o segundo menor recebe posto 2, e assim por diante.
4. Atribua sinais aos postos: se $D_i > 0$, mantenha o sinal positivo; se $D_i < 0$, atribua sinal negativo.

A estatística de Wilcoxon é definida como a soma dos postos positivos:

$$W^+ = \sum_{i:D_i>0} R_i \tag{189}$$

Alternativamente, podemos usar a soma dos postos negativos:

$$W^- = \sum_{i:D_i<0} R_i \tag{190}$$

Note que $W^+ + W^- = 1 + 2 + \dots + n' = \frac{n'(n'+1)}{2}$.

10.1.3 Distribuição da Estatística sob H_0

Teorema 4: Sob $H_0 : M = M_0$ e a suposição de simetria, a estatística W^+ tem distribuição que não depende da distribuição específica de X , mas apenas de n' .

A distribuição exata pode ser obtida enumerando todas as possíveis atribuições de sinais aos postos. Sob H_0 , cada D_i tem probabilidade $1/2$ de ser positivo (devido à simetria), independentemente.

Valor esperado e variância sob H_0 :

$$E_{H_0}[W^+] = \frac{n'(n'+1)}{4} \tag{191}$$

$$\text{Var}_{H_0}[W^+] = \frac{n'(n'+1)(2n'+1)}{24} \tag{192}$$

10.1.4 Distribuição Assintótica

Teorema 5: Sob H_0 e quando $n' \rightarrow \infty$, temos:

$$\frac{W^+ - E_{H_0}[W^+]}{\sqrt{\text{Var}_{H_0}[W^+]}} \xrightarrow{d} N(0, 1) \quad (193)$$

Demonstração (esboço): A estatística W^+ pode ser escrita como:

$$W^+ = \sum_{i=1}^{n'} R_i I_i \quad (194)$$

onde $I_i = 1$ se $D_i > 0$ e $I_i = 0$ caso contrário, e R_i são os postos.

Sob H_0 , os I_i são variáveis aleatórias independentes com $P(I_i = 1) = 1/2$. Os postos R_i são determinísticos (dependem apenas dos valores absolutos $|D_i|$).

Aplicando uma versão do Teorema Central do Limite para somas de variáveis independentes (não necessariamente identicamente distribuídas), obtemos o resultado assintótico.

Observação: Para amostras pequenas ($n' < 20$), é recomendado usar a distribuição exata. Para amostras maiores, a aproximação normal é adequada.

10.1.5 Regra de Decisão

Para um teste bilateral com nível α :

- **Amostras pequenas:** Rejeite H_0 se $W^+ \leq w_{\alpha/2}$ ou $W^+ \geq w_{1-\alpha/2}$, onde os valores críticos são obtidos da distribuição exata.
- **Amostras grandes:** Rejeite H_0 se

$$\left| \frac{W^+ - \frac{n'(n'+1)}{4}}{\sqrt{\frac{n'(n'+1)(2n'+1)}{24}}} \right| > z_{\alpha/2} \quad (195)$$

10.1.6 Exemplo Ilustrativo

Suponha que temos $n = 8$ observações e queremos testar $H_0 : M = 10$:

Dados: $X = \{12, 8, 15, 9, 11, 7, 13, 10\}$

Diferenças $D_i = X_i - 10$: $\{2, -2, 5, -1, 1, -3, 3, 0\}$

Removendo $D_8 = 0$, temos $n' = 7$.

Valores absolutos e postos:

$ D_i $	Posto R_i	Sinal	Posto com sinal
1	1	+	+1
1	2	-	-2
2	3	+	+3
2	4	-	-4
3	5	-	-5
3	6	+	+6
5	7	+	+7

$$W^+ = 1 + 3 + 6 + 7 = 17$$

$$W^- = 2 + 4 + 5 = 11$$

$$\text{Note que } W^+ + W^- = 28 = 7 \cdot 8/2.$$

10.2 Teste de Wilcoxon para Duas Amostras (Mann-Whitney-Wilcoxon)

Este teste é usado para comparar as medianas (ou mais geralmente, as distribuições) de duas populações independentes, sem assumir normalidade.

10.2.1 Hipóteses

Sejam X_1, \dots, X_n uma amostra de F_X e Y_1, \dots, Y_m uma amostra independente de F_Y , ambas contínuas. Queremos testar:

$$H_0 : F_X = F_Y \quad \text{vs} \quad H_1 : F_X \neq F_Y \quad (\text{ou } F_X(x) = F_Y(x - \Delta) \text{ com } \Delta \neq 0)$$

A hipótese alternativa de deslocamento ($F_X(x) = F_Y(x - \Delta)$) corresponde a $X \stackrel{d}{=} Y + \Delta$, ou seja, as distribuições são idênticas exceto por um deslocamento.

10.2.2 Construção da Estatística

O procedimento é:

1. Combine as duas amostras: $\{X_1, \dots, X_n, Y_1, \dots, Y_m\}$.
2. Ordene todas as $n + m$ observações do menor para o maior.
3. Atribua postos: a menor observação recebe posto 1, a segunda menor recebe posto 2, e assim por diante.
4. Em caso de empates, atribua o posto médio (average rank).
5. Seja R_i o posto de X_i na amostra combinada.

A estatística de Wilcoxon é a soma dos postos da primeira amostra:

$$W = \sum_{i=1}^n R_i \tag{196}$$

Observação: Esta estatística também é conhecida como estatística de Mann-Whitney, e há uma relação:

$$U = W - \frac{n(n+1)}{2} \tag{197}$$

onde U é a estatística de Mann-Whitney, que conta o número de pares (X_i, Y_j) tais que $X_i < Y_j$.

10.2.3 Distribuição da Estatística sob H_0

Teorema 6: Sob $H_0 : F_X = F_Y$, a distribuição de W não depende das distribuições específicas F_X e F_Y , mas apenas de n e m .

Valor esperado e variância sob H_0 :

$$E_{H_0}[W] = \frac{n(n+m+1)}{2} \tag{198}$$

$$\text{Var}_{H_0}[W] = \frac{nm(n+m+1)}{12} \quad (199)$$

Justificativa: Sob H_0 , todas as $n+m$ observações são provenientes da mesma distribuição. Portanto, os postos são uma permutação aleatória de $\{1, 2, \dots, n+m\}$. A soma dos postos da primeira amostra tem distribuição hipergeométrica.

10.2.4 Distribuição Assintótica

Teorema 7: Sob H_0 e quando $\min(n, m) \rightarrow \infty$ com $n/(n+m) \rightarrow \lambda \in (0, 1)$, temos:

$$\frac{W - E_{H_0}[W]}{\sqrt{\text{Var}_{H_0}[W]}} \xrightarrow{d} N(0, 1) \quad (200)$$

Demonstração (esboço): A estatística W pode ser escrita como:

$$W = \sum_{i=1}^n R_i = \sum_{i=1}^n \sum_{j=1}^{n+m} j \cdot I(X_i \text{ é a } j\text{-ésima menor observação}) \quad (201)$$

Sob H_0 , a distribuição dos postos é simétrica. Aplicando o Teorema Central do Limite para estatísticas de postos, obtemos a normalidade assintótica.

Correção de continuidade: Para amostras de tamanho moderado, uma correção de continuidade pode melhorar a aproximação:

$$Z = \frac{W - \frac{n(n+m+1)}{2} \pm 0.5}{\sqrt{\frac{nm(n+m+1)}{12}}} \quad (202)$$

onde o ± 0.5 é escolhido para aproximar W do valor mais próximo.

10.2.5 Regra de Decisão

Para um teste bilateral com nível α :

- **Amostras pequenas:** Use tabelas da distribuição exata de W (ou U).
- **Amostras grandes:** Rejeite H_0 se

$$\left| \frac{W - \frac{n(n+m+1)}{2}}{\sqrt{\frac{nm(n+m+1)}{12}}} \right| > z_{\alpha/2} \quad (203)$$

10.2.6 Tratamento de Empates

Quando há empates, a variância precisa ser ajustada. Seja t_j o número de observações empatadas no j -ésimo grupo de empates. A variância corrigida é:

$$\text{Var}_{H_0}[W] = \frac{nm(n+m+1)}{12} \left[1 - \frac{\sum_j (t_j^3 - t_j)}{(n+m)((n+m)^2 - 1)} \right] \quad (204)$$

10.2.7 Exemplo Ilustrativo

Suponha que temos duas amostras:

Amostra 1 ($n = 4$): $X = \{12, 15, 18, 20\}$

Amostra 2 ($m = 5$): $Y = \{10, 14, 16, 19, 22\}$

Amostra combinada ordenada: $\{10, 12, 14, 15, 16, 18, 19, 20, 22\}$

Postos:

Valor	Amostra	Posto
10	Y	1
12	X	2
14	Y	3
15	X	4
16	Y	5
18	X	6
19	Y	7
20	X	8
22	Y	9

$$W = 2 + 4 + 6 + 8 = 20$$

$$E_{H_0}[W] = \frac{4(4+5+1)}{12} = 20$$

$$\text{Var}_{H_0}[W] = \frac{4 \cdot \frac{2}{3} \cdot 10}{12} = \frac{200}{12} = 16.67$$

Neste caso, $W = E_{H_0}[W]$, então não há evidência contra H_0 .

10.3 Vantagens e Desvantagens dos Testes Não-Paramétricos

Vantagens:

- Não requerem suposições sobre a forma da distribuição (normalidade, etc.).
- São robustos a outliers.
- Podem ser mais poderosos que testes paramétricos quando as suposições paramétricas não são satisfeitas.
- Aplicáveis a dados ordinais (não apenas contínuos).

Desvantagens:

- Podem ser menos poderosos que testes paramétricos quando as suposições paramétricas são satisfeitas.
- Para amostras pequenas, podem requerer tabelas especiais.
- A interpretação pode ser menos intuitiva (testam medianas ou distribuições, não médias).

11 Resumo e Observações Finais

Este material auxiliar apresentou de forma organizada os principais conceitos e métodos de inferência estatística relacionados a intervalos de confiança e testes de hipóteses baseados na razão de verossimilhanças.

11.1 Principais Conceitos Abordados

- **Intervalos de Confiança:** Duas abordagens principais foram apresentadas: inversão de testes de hipóteses e uso de quantidades pivotais. Ambas são equivalentes em muitos casos, mas a abordagem pivotal é frequentemente mais direta.
- **Teste de Razão de Verossimilhanças:** Método geral para construção de testes de hipóteses que não requer a existência de testes UMP. É particularmente útil para testes bilaterais e casos com múltiplos parâmetros.
- **Testes Assintóticos:** Os testes de Wald e Escore fornecem alternativas assintoticamente equivalentes ao TRV, baseadas em propriedades assintóticas do EMV. São especialmente úteis quando a distribuição exata é difícil de obter.
- **Testes Não-Paramétricos:** O teste de Wilcoxon oferece uma alternativa robusta aos testes paramétricos, não requerendo suposições sobre a forma da distribuição. Inclui versões para uma e duas amostras.
- **Distribuições Normais:** A maior parte dos exemplos considerou populações normais, pois permitem derivações exatas e são amplamente aplicáveis na prática.
- **Comparação de Populações:** Métodos para comparar parâmetros de duas populações normais independentes, incluindo comparação de médias e variâncias.

11.2 Observações Importantes

1. A escolha entre usar distribuição normal ou t de Student depende do conhecimento da variância populacional. Quando a variância é desconhecida, sempre use a distribuição t .
2. Para intervalos de confiança de variâncias, a distribuição qui-quadrado não é simétrica, então os quantis $\chi_{\alpha/2}^2$ e $\chi_{1-\alpha/2}^2$ não são simétricos.
3. Em testes para duas amostras, a suposição de variâncias iguais é crucial para alguns métodos. Quando essa suposição não é válida, métodos alternativos devem ser considerados.
4. A interpretação de intervalos de confiança e testes de hipóteses é diferente: intervalos fornecem uma faixa de valores plausíveis para o parâmetro, enquanto testes avaliam evidência contra uma hipótese específica.
5. Os testes de Wald, Escore e TRV são assintoticamente equivalentes, mas podem diferir em amostras finitas. A escolha entre eles pode depender de considerações computacionais e de invariância.
6. Testes não-paramétricos como o de Wilcoxon são particularmente valiosos quando há dúvidas sobre suposições distribucionais ou quando se trabalha com dados ordinais. No entanto, podem ser menos poderosos que testes paramétricos quando as suposições paramétricas são satisfeitas.
7. Para amostras pequenas, é recomendado usar distribuições exatas quando disponíveis. Aproximações assintóticas devem ser usadas com cautela para $n < 30$ em muitos casos.

Este material foi compilado a partir das notas de aula do curso de Inferência Estatística, preservando a notação e o estilo do professor.