Estatística não paramétrica: Métodos baseados em postos

Dorival Leão

Portal Action

2021

Técnicas não paramétricas: Uso

- Quando as hipóteses a serem testadas não envolvem parâmetros;
- Os dados são medidas em uma escala de discriminação ruim para aplicar métodos paramétricos. Por exemplo, se os dados são contagens ou postos;
- As suposições necessárias para validar um procedimento paramétrico não são válidas;
- Necessidade de modelagem mais flexível e menos restritiva.

Técnicas não paramétricas: Uso

- Quando as hipóteses a serem testadas não envolvem parâmetros;
- Os dados são medidas em uma escala de discriminação ruim para aplicar métodos paramétricos. Por exemplo, se os dados são contagens ou postos;
- As suposições necessárias para validar um procedimento paramétrico não são válidas;
- Necessidade de modelagem mais flexível e menos restritiva.

Técnicas não paramétricas: Uso

- Quando as hipóteses a serem testadas não envolvem parâmetros;
- Os dados são medidas em uma escala de discriminação ruim para aplicar métodos paramétricos. Por exemplo, se os dados são contagens ou postos;
- As suposições necessárias para validar um procedimento paramétrico não são válidas;
- Necessidade de modelagem mais flexível e menos restritiva.

Técnicas não paramétricas: Uso

- Quando as hipóteses a serem testadas não envolvem parâmetros;
- Os dados são medidas em uma escala de discriminação ruim para aplicar métodos paramétricos. Por exemplo, se os dados são contagens ou postos;
- As suposições necessárias para validar um procedimento paramétrico não são válidas;
- Necessidade de modelagem mais flexível e menos restritiva.

Teste de Wilcoxon

- Consideremos uma população $\mathbb{P}$ com distribuição contínua e simétrica no qual retiramos uma amostra aleatória smples $X_1, \cdots, X_n$;

- O teste de Wilcoxon é baseado nos postos (ranks) dos valores obtidos. Postos são as posições, representados por números, que os valores obervados ocupam quando colocados em ordem crescente. Por exemplo, considere o seguinte conjunto de valores: $12, 17, 15, 19, 14, 16$ e 11;

- Colocando em ordem crescente e atribuindo a cada valor seu posto, temos

Valor	11	12	14	15	16	17	19
posto	1	2	3	4	5	6	7

Teste de Wilcoxon

- Consideremos uma população $\mathbb{P}$ com distribuição contínua e simétrica no qual retiramos uma amostra aleatória smples $X_1, \cdots, X_n$;

- O teste de Wilcoxon é baseado nos postos (ranks) dos valores obtidos. Postos são as posições, representados por números, que os valores obervados ocupam quando colocados em ordem crescente. Por exemplo, considere o seguinte conjunto de valores: $12, 17, 15, 19, 14, 16$ e 11;

- Colocando em ordem crescente e atribuindo a cada valor seu posto, temos

Valor	11	12	14	15	16	17	19
posto	1	2	3	4	5	6	7

Teste de Wilcoxon

- Consideremos uma população $\mathbb{P}$ com distribuição contínua e simétrica no qual retiramos uma amostra aleatória smples $X_1, \cdots, X_n$;

- O teste de Wilcoxon é baseado nos postos (ranks) dos valores obtidos. Postos são as posições, representados por números, que os valores obervados ocupam quando colocados em ordem crescente. Por exemplo, considere o seguinte conjunto de valores: $12, 17, 15, 19, 14, 16$ e 11;

- Colocando em ordem crescente e atribuindo a cada valor seu posto, temos

Valor	11	12	14	15	16	17	19
posto	1	2	3	4	5	6	7

Hipóteses:

- As observações Xi's são independentes;
- Cada observação Xi é obtida de uma população que é contínua e simétrica em torno de θ_0. Desta forma, admitimos que a probabilidade (em teoria) de que dois valores amostrais coincidam é zero;

Para a realização do teste de Wilcoxon estabelecemos uma das seguintes hipóteses:

$$\begin{cases} H_0 : \theta - \theta_0 = 0 \\ H_1 : \theta - \theta_0 \neq 0 \end{cases} \quad \begin{cases} H_0 : \theta - \theta_0 = 0 \\ H_1 : \theta - \theta_0 > 0 \end{cases} \quad \begin{cases} H_0 : \theta - \theta_0 = 0 \\ H_1 : \theta - \theta_0 < 0 \end{cases}$$

Hipóteses:

- As observações Xi's são independentes;
- Cada observação Xi é obtida de uma população que é contínua e simétrica em torno de θ_0. Desta forma, admitimos que a probabilidade (em teoria) de que dois valores amostrais coincidam é zero;

Para a realização do teste de Wilcoxon estabelecemos uma das seguintes hipóteses:

$$\begin{cases} H_0: \theta - \theta_0 = 0 \\ H_1: \theta - \theta_0 \neq 0 \end{cases} \quad \begin{cases} H_0: \theta - \theta_0 = 0 \\ H_1: \theta - \theta_0 > 0 \end{cases} \quad \begin{cases} H_0: \theta - \theta_0 = 0 \\ H_1: \theta - \theta_0 < 0 \end{cases}$$

Teste de Wilcoxon

- Subtrair θ_0 de cada valor $X_1, \cdots, X_N$ da amostra e assim obtemos um novo conjunto de dados: $Z_1, Z_2, \cdots, Z_n$ no qual $Z_i = X_i - \theta_0$;

- Ordenamos de forma crescente o novo conjunto de dados $\{| Z_1 |, | Z_2 |, \cdots, | Z_n |\}$ e associamos a cada valor Z_i o posto R_i correspondente. A seguir, definimos a variável indicadora Ψ_i na forma

$$\psi_i = \begin{cases} 1, & \text{se } Z_i > 0 \\ 0, & \text{se } Z_i < 0 \end{cases},$$

para todo $i = 1, \cdots, n$.

- Definimos a estatística T^+ como a soma dos postos que têm sinal positivo, ou seja,

$$T^+ = \sum_{i}^{n} R_i \psi_i$$

Teste de Wilcoxon

- Subtrair θ_0 de cada valor $X_1, \cdots, X_N$ da amostra e assim obtemos um novo conjunto de dados: $Z_1, Z_2, \cdots, Z_n$ no qual $Z_i = X_i - \theta_0$;

- Ordenamos de forma crescente o novo conjunto de dados $\{| Z_1 |, | Z_2 |, \cdots, | Z_n |\}$ e associamos a cada valor Z_i o posto R_i correspondente. A seguir, definimos a variável indicadora Ψ_i na forma

$$\psi_i = \begin{cases} 1, \text{ se } Z_i > 0 \\ 0, \text{ se } Z_i < 0 \end{cases},$$

para todo $i = 1, \cdots, n$.

- Definimos a estatística T^+ como a soma dos postos que têm sinal positivo, ou seja,

$$T^+ = \sum_i^n R_i \psi_i$$

Teste de Wilcoxon

- Subtrair θ_0 de cada valor $X_1, \cdots, X_N$ da amostra e assim obtemos um novo conjunto de dados: $Z_1, Z_2, \cdots, Z_n$ no qual $Z_i = X_i - \theta_0$;

- Ordenamos de forma crescente o novo conjunto de dados $\{| Z_1 |, | Z_2 |, \cdots, | Z_n |\}$ e associamos a cada valor Z_i o posto R_i correspondente. A seguir, definimos a variável indicadora Ψ_i na forma

$$\psi_i = \begin{cases} 1, \text{ se } Z_i > 0 \\ 0, \text{ se } Z_i < 0 \end{cases},$$

para todo $i = 1, \cdots, n$.

- Definimos a estatística T^+ como a soma dos postos que têm sinal positivo, ou seja,

$$T^+ = \sum_i^n R_i \psi_i$$

Distribuição Exata

Sob H0 as distribuições de todos os $Z_1, \cdots, Z_n$ são simétricas em torno de $\theta = \theta_0$. Portanto, se temos uma amostra de n elementos, temos 2^n possibilidades para a configuração $(R_1, R_2, \cdots, R_B)$ e cada uma delas ocorre com probabilidade $\frac{1}{2^n}$, no qual B é o número de postos com Z_i positivo. Neste caso, temos que

$$P(T^+ = t) = \frac{u(t)}{2^n};$$

no qual $u(t)$ é o número de maneiras de atribuir valores para as configurações $(R_1, R_2, \cdots, R_B)$ de forma que

$$\sum_{i=1}^{B} R_i = t.$$

Distribuição Assintótica

Sob a hipótese nula ($\theta = \theta_0$), temos que o valor esperado e a variância são dados por

$$E(T^+ \mid \theta = \theta_0) = \frac{n(n+1)}{4} \quad \text{e} \quad Var(T^+ \mid \theta = \theta_0) = \frac{n(n+1)(2n+1)}{24}.$$

A estatística Z, dada por

$$Z = \frac{T^+ - E_0(T^+)}{\sqrt{Var_0(T^+)}} = \frac{T^+ - \frac{1}{4}n(n+1)}{\sqrt{[n(n+1)(2n+1)/24]}}.$$

tem distribuição aproximadamente Normal com média 0 e variância 1.

Distribuição Assintótica

Sob a hipótese nula ($\theta = \theta_0$), temos que o valor esperado e a variância são dados por

$$E(T^+ \mid \theta = \theta_0) = \frac{n(n+1)}{4} \quad \text{e} \quad Var(T^+ \mid \theta = \theta_0) = \frac{n(n+1)(2n+1)}{24}.$$

A estatística Z, dada por

$$Z = \frac{T^+ - E_0(T^+)}{\sqrt{Var_0(T^+)}} = \frac{T^+ - \frac{1}{4}n(n+1)}{\sqrt{[n(n+1)(2n+1)/24]}}.$$

tem distribuição aproximadamente Normal com média 0 e variância 1.

Estatística de Postos: Processo Pontual

- Seja X uma variável aleatória discreta assumindo valores $\{1, 2, \cdots, k\}$ no qual k pode ser finito ou não;
- Tomamos $N(i) = \mathbb{1}_{\{X \leq i\}}$ para todo $i = 1, 2, \cdots$. Neste caso, obtemos que N é um submartingale;
- Decomposição de Doob nos garante que $N = Y + A$, no qual Y é um martingale e A é um processo não decrescente tal que

$$A(i) = \sum_{\ell=1}^{i} h(\ell) V(\ell),$$

no qual

$$h(\ell) = \frac{\mathbb{P}[X = \ell]}{\mathbb{P}[X \geq \ell]} \quad \text{e} \quad V(\ell) = \mathbb{1}_{\{X \geq \ell\}}.$$

Estatística de Postos: Função Intensidade

- Dado X uma variável aleatória discreta com distribuição de probabilidade $P[X = i]$. Então, existe uma única função intensidade $h : \{1, 2, \cdots, k\} \to [0, 1]$, tal que

$$h(i) \;=\; \frac{P[X = i]}{P[X \geq i]},$$

com $i = 1, \cdots, k$.

- Dado uma função de intensidade h, existe uma única variável aleatória X satisfazendo

$$\mathbb{P}[X = i] \;=\; h(i)\prod_{\ell=1}^{i-1}[1 - h(\ell)] = \prod_{\ell=1}^{k}[1 - h(\ell)]^{V(\ell)-\Delta N(\ell)}[h(\ell)]^{\Delta N(\ell)},$$

na qual $V(\ell) = 1\!\!1_{\{X \geq \ell\}}$ e $N(\ell) = 1\!\!1_{\{X \leq \ell\}}$, para todo $i = 1, \cdots, k$.

Estatística de Postos: Modelo

- Tomamos $X_1^p, \cdots, X_{n_p}^p$ uma amostra aleatória simples (iid) da distribuição F^p, tal que $n_p \geq 1$ e $p = 1, \cdots, J$;
- Assumimos que as populações também são independentes;
- Como as variáveis aleatórias são independentes, temos que

$$N^{n_p}(i) = \sum_{m=1}^{n_p} \mathbb{1}_{\{X_m^p \leq i\}} = Y^p(i) + A^{n_p}(i),$$

nos quais

$$A^{n_p}(i) = \sum_{\ell=1}^{i} V^{n_p}(\ell) h^p(\ell) \quad \text{e} \quad V^{n_p}(\ell) = \sum_{m=1}^{n_p} \mathbb{1}_{\{X_m^p \leq \ell\}}$$

Estatística de Postos: Kaplan-Meier

- Na decomposição de Doob, o martingale Y^{n_p} representa o ruído. Assim, queremos um estimador que elimine o ruído. Neste caso, queremos encontrar $\hat{h}^{n_p}$ tal que $Y^{n_p}(i) = 0$ para todo $i = 1, 2, \cdots, k$;
- Temos que $Y^{n_p}(1) = 0$ implica que

$$\hat{h}^{n_p}(1) = \frac{N^{n_p}(1)}{V^{n_p}(1)} = \frac{\Delta N^{n_p}(1)}{V^{n_p}(1)}, \quad N^{n_p}(0) = 0;$$

- Da mesma forma, se $Y^{n_p}(\ell) = 0$ obtemos o estimador de Kaplan-Meier

$$\hat{h}^{n_p}(\ell) = \frac{\Delta N^{n_p}(\ell)}{V^{n_p}(\ell)}, \quad \Delta N^{n_p}(\ell) = N^{n_p}(\ell) - N^{n_p}(\ell - 1),$$

para todo $\ell = 1, 2, \cdots, k$;

Estatística de Postos: Kaplan-Meier

- Como V^{n_p} é um processo previsível concluímos que

$$\sum_{\ell=1}^{i} \hat{h}^{n_p}(\ell) = \underbrace{\sum_{\ell=1}^{i} \frac{1}{V^{n_p}(\ell)} \Delta Y^{n_p}(\ell)}_{martingale} + \sum_{\ell=1}^{i} h^p(\ell);$$

- Assim, obtemos que

$$\mathbb{E}\hat{h}^{n_p}(i) = h^p(i), \quad i = 1, 2, \cdots, k;$$

- Temos que

$$\begin{aligned} \widehat{\mathbb{P}[x = i]} &= \prod_{\ell=1}^{i} [1 - \hat{h}^{n_p}(\ell)]^{V^{n_p}(\ell) - \Delta N^{n_p}(\ell)} [\hat{h}^{n_p}(\ell)]^{V^{n_p}(\ell)} \\ &= \frac{\Delta N^{n_p}(i)}{n_p}, \quad \text{Funcão de Distribuição Empírica.} \end{aligned}$$

Estatística de Postos: Kaplan-Meier

- Como V^{n_p} é um processo previsível concluímos que

$$\sum_{\ell=1}^{i} \hat{h}^{n_p}(\ell) = \underbrace{\sum_{\ell=1}^{i} \frac{1}{V^{n_p}(\ell)} \Delta Y^{n_p}(\ell)}_{\textit{martingale}} + \sum_{\ell=1}^{i} h^p(\ell);$$

- Assim, obtemos que

$$\mathbb{E}\hat{h}^{n_p}(i) = h^p(i), \quad i = 1, 2, \cdots, k;$$

- Temos que

$$\begin{aligned} \widehat{\mathbb{P}[x = i]} &= \prod_{\ell=1}^{i} [1 - \hat{h}^{n_p}(\ell)]^{V^{n_p}(\ell) - \Delta N^{n_p}(\ell)} [\hat{h}^{n_p}(\ell)]^{V^{n_p}(\ell)} \\ &= \frac{\Delta N^{n_p}(i)}{n_p}, \quad \text{Funcão de Distribuição Empírica.} \end{aligned}$$

Estatística de Postos: Kaplan-Meier

- Como V^{n_p} é um processo previsível concluímos que

$$\sum_{\ell=1}^{i} \hat{h}^{n_p}(\ell) = \underbrace{\sum_{\ell=1}^{i} \frac{1}{V^{n_p}(\ell)} \Delta Y^{n_p}(\ell)}_{martingale} + \sum_{\ell=1}^{i} h^p(\ell);$$

- Assim, obtemos que

$$\mathbb{E}\hat{h}^{n_p}(i) = h^p(i), \quad i = 1, 2, \cdots, k;$$

- Temos que

$$\begin{aligned} \widehat{\mathbb{P}[x = i]} &= \prod_{\ell=1}^{i} [1 - \hat{h}^{n_p}(\ell)]^{V^{n_p}(\ell) - \Delta N^{n_p}(\ell)} [\hat{h}^{n_p}(\ell)]^{V^{n_p}(\ell)} \\ &= \frac{\Delta N^{n_p}(i)}{n_p}, \quad \text{Funcão de Distribuição Empírica.} \end{aligned}$$

Estatística de Postos: Kaplan-Meier

- A variância do estimador de Kaplan-Meier é dado por

$$Var[\hat{h}^{n_p}(\ell)] = h^{(p)}(\ell)\left[1 - h^{(p)}(\ell)\right]\mathbb{E}\left[\frac{1}{V^{n_p}(\ell)}\right],\ \ell \leq i;$$

- Temos que $V^{n_p}(\ell)$ tem distribuição Binomial com parâmetros n_p e $\mathbb{P}[X^p \geq \ell]$;

Estatística de Postos: Distribuição Assintótica

- Dada uma sequência $\{\hat{h}^{n_p}\}$ de estimadores da função intensidade $h^{(p)}$ temos que
$$\mathbb{P}\left[\sup_{\ell \leq i} |\hat{h}^{n_p}(\ell) - h^{(p)}(\ell)|\right] \to 0,$$
ou seja, $\hat{h}^{n_p}$ é um estimador consistente para a função $h^{(p)}$;

- O estimador $\widehat{\mathbb{P}[X = i]}$ é consistente para da função de distribuição $\mathbb{P}[X = i]$;

- Temos que
$$\sqrt{n_p}\left[\hat{h}^p(\ell) - h^p(\ell)\right]$$
converge em distribuição para uma variável aleatória normal com média zero e variância
$$\sigma_h^2(\ell) = \frac{h^p(\ell)\,(1 - h^p(\ell))}{\mathbb{P}[X^p \geq \ell]}, \quad \ell = 1, 2, \cdots, k.$$

Estatística de Postos: Distribuição Assintótica

- Dada uma sequência $\{\hat{h}^{n_p}\}$ de estimadores da função intensidade $h^{(p)}$ temos que

$$\mathbb{P}\left[\sup_{\ell \leq i} |\hat{h}^{n_p}(\ell) - h^{(p)}(\ell)|\right] \to 0,$$

ou seja, $\hat{h}^{n_p}$ é um estimador consistente para a função $h^{(p)}$;

- O estimador $\widehat{\mathbb{P}[X = i]}$ é consistente para da função de distribuição $\mathbb{P}[X = i]$;

- Temos que

$$\sqrt{n_p}\left[\hat{h}^p(\ell) - h^p(\ell)\right]$$

converge em distribuição para uma variável aleatória normal com média zero e variância

$$\sigma_h^2(\ell) = \frac{h^p(\ell)\,(1 - h^p(\ell))}{\mathbb{P}[X^p \geq \ell]}, \quad \ell = 1, 2, \cdots, k.$$

Estatística de Postos: Distribuição Assintótica

- Dada uma sequência $\{\hat{h}^{n_p}\}$ de estimadores da função intensidade $h^{(p)}$ temos que
$$\mathbb{P}\left[\sup_{\ell \leq i} |\hat{h}^{n_p}(\ell) - h^{(p)}(\ell)|\right] \to 0,$$
ou seja, $\hat{h}^{n_p}$ é um estimador consistente para a função $h^{(p)}$;

- O estimador $\widehat{\mathbb{P}[X = i]}$ é consistente para da função de distribuição $\mathbb{P}[X = i]$;
- Temos que
$$\sqrt{n_p}\left[\hat{h}^p(\ell) - h^p(\ell)\right]$$
converge em distribuição para uma variável aleatória normal com média zero e variância
$$\sigma_h^2(\ell) = \frac{h^p(\ell)\,(1 - h^p(\ell))}{\mathbb{P}[X^p \geq \ell]}, \quad \ell = 1, 2, \cdots, k.$$

Teste de Logrank Ponderado

- Denotamos $n^\star := (n_1, \ldots, n_J) \in \mathbb{N}^J$, $N_J := \max\{n_1, \ldots, n_J\}$ e $n = n_1 + \cdots + n_J$;
- $n^\star \to \infty$ significa $n_p \to \infty$ para todo $p \in \mathcal{J}$ e assumimos que $b_p := \lim_{n^\star \to \infty} n_p/n; p \in \mathcal{J}$ existe;
- Queremos testar as hipóteses

$$H_0 : h^1(\ell) = h^2(\ell) = \cdots = h^J(\ell); \ell \in \mathbb{N};$$

$$H_1 : \text{Pelos menos uma função intensidade diferente};$$

- Seja $R^{n^\star}(i) := \sum_{p=1}^{J} R^{n_p}(i)$ e $V^{n^\star}(i) = \sum_{p=1}^{J} V^{n_p}(i)$ o número total de eventos de interesse e o número total sob risco na categoria i, respectivamente.

Estatística de Logrank Ponderado

Estatística de Logrank ponderado é dada por:

$$
\begin{aligned}
LR_q(n^\star, i) \quad &:= \quad \sum_{\ell=1}^{i} \sum_{q_1 \neq q} \left(\frac{1}{n}\right)^{1/2} u(n^\star, \ell) \left(\frac{V^{n_q}(\ell) V^{n_{q_1}}(\ell)}{V^{n^\star}(\ell)} \right) \left[\hat{h}^{n_q}(\ell) - \hat{h}^{n_{q_1}}(\ell) \right] \\
&= \quad \sum_{\ell=1}^{i} \left(\frac{1}{n}\right)^{1/2} u(n^\star, \ell) \left[\Delta R^{n_q}(\ell) - V^{n_q}(\ell) \frac{\Delta R^{n^\star}(\ell)}{V^{n^\star}(\ell)} \right], \quad q \in \mathcal{J},
\end{aligned}
$$

no qual o processo peso $u(n^\star; \cdot)$ é limitado e converge em probabilidade para uma função limitada.

Estatística de Logrank Ponderado: Pesos

- Gehan e Wilcoxon ($J = 2$) e Kruskal-Walis ($J > 2$)

$$u(n^\star, \ell) := \frac{V^{n^\star}(\ell)}{n}$$

- Tarone e Ware

$$u(n^\star, \ell) := \left(\frac{V^{n^\star}(\ell)}{n}\right)^\gamma, \quad \gamma > 0;$$

- Logrank

$$u(n^\star, \ell) := 1.$$

Estatística de Logrank Ponderado: Pesos

- Gehan e Wilcoxon ($J = 2$) e Kruskal-Walis ($J > 2$)

$$u(n^\star, \ell) := \frac{V^{n^\star}(\ell)}{n}$$

- Tarone e Ware

$$u(n^\star, \ell) := \left(\frac{V^{n^\star}(\ell)}{n}\right)^\gamma, \quad \gamma > 0;$$

- Logrank

$$u(n^\star, \ell) := 1.$$

Estatística de Logrank Ponderado: Pesos

Como a soma dos componentes LR_q é nula, a estatística de Logrank multivariada é definida por:

$$LR_0(n^\star, i) := (LR_1(n^\star, i), \ldots, LR_{J-1}(n^\star, i))^\top,$$

para $n^\star \in \mathbb{N}^J$ e $i \geq 1$. A estrutura de covariância é dada por

$$\left[\hat{\Gamma}(n^\star, i)\right]_{rk} := \begin{cases} \sum_{\ell=1}^{i} \hat{\phi}^2_{k,n^\star}(\ell); & \text{if } k = r \\ \sum_{\ell=1}^{i} \hat{\psi}_{n^\star}(k, r, \ell); & \text{if } k \neq r, \end{cases} \quad (1)$$

Estatística de Logrank Ponderado: Pesos

Variância:

$$\hat{\phi}^2_{q,n^\star}(\ell) \quad = \quad \frac{1}{n}u^2(n^\star,\ell)\frac{V^{n_q}(\ell)}{V^{n^\star(\ell)}(\ell)}\left[1-\frac{V^{n_q}(\ell)}{V^{n^\star}(\ell)}\right]\left[\frac{V^{n^\star}(\ell)-\Delta R^{n^\star}(\ell)}{V^{n^\star}(\ell)-1}\right]\Delta R^{n^\star}(\ell).$$

Covariância:

$$\hat{\psi}_{n^\star}(k,r,\ell) \quad = \quad -\frac{1}{n}u^2(n^\star,\ell)\frac{V^{n_k}(\ell)}{V^{n^\star(\ell)}(\ell)}\frac{V^{n_r}(\ell)}{V^{n^\star(\ell)}(\ell)}\left[\frac{V^{n^\star}(\ell)-\Delta R^{n^\star}(\ell)}{V^{n^\star}(\ell)-1}\right]\Delta R^{n^\star}(\ell).$$

se $r \neq k$ em $\mathcal{J}$.

Estatística de Logrank Ponderado: Pesos

Variância:

$$\hat{\phi}^2_{q,n^\star}(\ell) \quad = \quad \frac{1}{n}u^2(n^\star,\ell)\frac{V^{n_q}(\ell)}{V^{n^\star(\ell)}(\ell)}\left[1-\frac{V^{n_q}(\ell)}{V^{n^\star}(\ell)}\right]\left[\frac{V^{n^\star}(\ell)-\Delta R^{n^\star}(\ell)}{V^{n^\star}(\ell)-1}\right]\Delta R^{n^\star}(\ell).$$

Covariância:

$$\hat{\psi}_{n^\star}(k,r,\ell) \quad = \quad -\frac{1}{n}u^2(n^\star,\ell)\frac{V^{n_k}(\ell)}{V^{n^\star(\ell)}(\ell)}\frac{V^{n_r}(\ell)}{V^{n^\star(\ell)}(\ell)}\left[\frac{V^{n^\star}(\ell)-\Delta R^{n^\star}(\ell)}{V^{n^\star}(\ell)-1}\right]\Delta R^{n^\star}(\ell).$$

se $r \neq k$ em $\mathcal{J}$.

Estatística de Logrank Ponderado: Pesos

- Sob H_0, o vetor aleatório $LR_0(n^\star, i)$ converge em distribuição para uma distribuição normal multivariado $N\big(0, \Gamma_0(i)\big)$ quando $n^\star \to \infty$, no qual a matriz de covariâncias $\Gamma_0(i)$ pode ser consistentemente estimada por $\hat{\Gamma}_0(n^\star, i)$, no qual $\hat{\Gamma}_0(n^\star, \cdot)$ é dada por $\hat{\Gamma}(n^\star, \cdot)$ sem a última linha e a última coluna;

- A estatística de logrank ponderada relacionada com LR_0 dada por

$$X^2(n^\star, i) := LR_0(n^\star, i)^\top \hat{\Gamma}_0(n^\star, i)^{-1} LR_0(n^\star, i), \ n^\star \in \mathbb{N}^J,$$

converge em distribuição para um chi-quadrado $J - 1$-graus de liberdade.