

Modelagem Conjunta de Dados Longitudinais e de Sobrevivência

Anna Carolina Lustosa Lima

1. Introdução

Em estudos envolvendo análise de dados longitudinais a variável resposta é medida repetidamente, ao longo de um intervalo de tempo, para cada indivíduo. Nos estudos de análise de sobrevivência a variável resposta é o tempo até a ocorrência de um evento. Muitos estudos clínicos geram conjuntamente dados longitudinais e de sobrevivência. Por exemplo, nos estudos prospectivos de análise de sobrevivência em que, grupos de pacientes são acompanhados ao longo do tempo até a ocorrência de óbito, é comum haver visitas de acompanhamento quando são medidas variáveis relacionadas à progressão da doença. Entretanto, tratar estas medidas como covariáveis dependentes do tempo no processo de sobrevivência nem sempre é satisfatório uma vez que há ocasiões cujo interesse do pesquisador é avaliar o efeito de tratamentos tanto sobre os tempos de sobrevivência quanto sobre o marcador longitudinal. Este estudo mostra também que a inclusão do marcador longitudinal como covariável no processo de sobrevivência pode ocasionar a perda de importância de covariáveis possivelmente associadas ao tratamento.

Existem modelos bem estabelecidos na literatura para analisar estes dados separadamente, mas esta forma de abordagem pode ser inadequada se a variável longitudinal for correlacionada com o tempo de sobrevivência. Por exemplo, Henserson et.al. (2000) mostraram que, em detrimento do modelo conjunto, utilizar os modelos separados para o processo longitudinal e de sobrevivência acarreta um vício severo na estimação de alguns componentes de variância e no coeficiente de tendência temporal. Guo e Carlin (2004) desenvolveram uma aproximação Bayesiana para o modelo conjunto e mostraram haver significativa melhora nas estimativas dos tempos de sobrevivência e de outros parâmetros.

O presente estudo pretende comparar as abordagens, separada e conjunta, aplicada ao banco de dados de um experimento biológico envolvendo camundongos, realizado com o objetivo de avaliar o efeito da co-infecção de malária e esquistossomose e da imunização induzida para malária, sobre os

tempos de sobrevivência e nível de parasitemia (malária) dos camundongos. Neste experimento, 64 camundongos foram distribuídos aleatoriamente em quatro grupos experimentais. Os camundongos alocados no primeiro e segundo grupos foram ambos previamente induzidos à imunização para malária e infectados com malária e malária associada à esquistossomose, respectivamente. Os camundongos alocados nos grupos 3 e 4 não foram imunizados e foram infectados por malária e malária associada à esquistossomose, respectivamente. O pesquisador visitou os camundongos diariamente e registrou o percentual médio de parasitemia no sangue e a ocorrência ou não de morte. Esta avaliação foi realizada num período de 25 dias (do 5º ao 30º dia após a infecção). Observou-se que os camundongos dos grupos 1 e 2 (ambos imunizados) tiveram melhor prognóstico e níveis de parasitemia baixos, enquanto que nos grupos 3 e 4 foram observados menores tempos de sobrevivência e níveis de parasitemia crescentes.

Em específico pretende-se comparar em termos metodológicos e de interpretação os seguintes modelos:

M1: Modelos para dados longitudinais;

M2: Modelo de Cox para a resposta de sobrevivência;

M3: Modelo de Cox com covariável dependente do tempo (variável longitudinal);

M4: Modelo Conjunto para dados longitudinais e de sobrevivência.

A análise conjunta de dados longitudinais de sobrevivência tem sido explorada recentemente por muitos autores Wulfshon e Tsiatis (1997) desenvolveram uma aproximação para o modelo conjunto que utiliza o modelo de Cox para o processo de sobrevivência e a resposta longitudinal é tratada como covariável contínua cujo valor real é dado por um modelo linear de efeitos aleatórios. A estimação dos parâmetros do modelo proposto é realizada por um algoritmo EM que também foi utilizado por Wenderson et. al. (2000) estes autores formularam um modelo generalizado para a análise conjunta em que um sub-modelo linear de efeitos aleatórios é assumido para os dados longitudinais, um sub-modelo semi-paramétrico de riscos proporcionais para os dados de sobrevivência, e assume que estes dois sub-modelos podem ser ligados por um processo latente normal bivariado. Uma característica interessante do modelo proposto por estes autores é que na ausência de associação entre o marcador longitudinal e os tempos de sobrevivência, os

resultados da análise conjunta deverão ser os mesmos que obteríamos a partir da análise separada de cada um dos componentes através de métodos tradicionais.

Outros autores também utilizaram esta forma generalizada para o modelo conjunto. Por exemplo, Wang e Taylor (2001) desenvolveram uma aproximação que considera o processo latente de ligação como um processo estocástico integrado Ornstein-Uhlenbeck. Esta escolha foi motivada pela necessidade de obter um modelo mais flexível para o processo longitudinal. Brown, Ibrahim e DeGruttola (2005) também sugeriram uma alteração no modelo para a resposta longitudinal. Eles sugerem o uso do modelo B-Splines Cúbico em lugar de um modelo linear de efeitos aleatórios.

Song, Davidian e Tsiatis (2002) avaliaram o modelo conjunto assumindo apenas que os coeficientes aleatórios pertencem a uma família de densidades suavizadas e não estritamente a distribuição normal. Guo e Carlin (2004) desenvolveram uma aproximação Bayesiana para o modelo conjunto proposto por Henderson et. al. (2000).

Tseng, Hsieh e Wang (2005) abordaram o caso em que a suposição de riscos proporcionais no modelo de Cox não é verificada e considera o modelo de tempo de tempos de falha acelerado como alternativa para a modelagem dos dados de eventos.

Ratcliffe et. al. (2004) propuseram um modelo conjunto para dados em conglomerados em que os dois sub-modelos são ligados através das fragilidades no nível do grupo.

Zhang e Heagerty (2005) e Xu e Zeng (2001) desenvolveram aproximações de modelagem conjunta mas, com o foco na estimação dos parâmetros do processo de sobrevivência.

A Seção 2 apresenta de forma resumida os conceitos de análise de sobrevivência e o modelo de Cox (1972) com e sem covariáveis dependentes do tempo. A Seção 3 aborda a metodologia de análise de dados longitudinais e apresenta os modelos marginal e de efeitos aleatórios. Os modelos conjuntos propostos por Wulfsohn e Tsiatis (1997) e por Henderson et. al. (2000) são apresentados na Seção 4. A descrição detalhada do experimento e modelagem dos dados utilizando os métodos tradicionais de análise separada estão apresentados na Seção 5. A Seção 6 considera o modelo de Cox com covariável dependente do tempo como uma aproximação para o modelo conjunto.

2. Análise de Sobrevida

Nesta seção introduziremos os conceitos básicos de análise de sobrevivência e a notação utilizada na literatura. Os modelos de Cox, com e sem covariável dependente do tempo, e suas formas de estimação são também apresentados nesta seção.

Suponha um conjunto de m indivíduos distribuídos aleatoriamente em grupos de tratamento, X_i é a variável indicadora de grupo. Seja T_i o tempo máximo de acompanhamento do i -ésimo indivíduo. Este tempo pode ser classificado de duas formas: T_i é um **tempo de falha** se indica o tempo em que o evento de interesse foi observado; ou, T_i é um **tempo de censura** se ao final do acompanhamento o evento não ocorreu.

De forma geral, o vetor (T_i, δ_i, X_i) representa os dados de sobrevivência do i -ésimo indivíduo em que t_i é o tempo máximo de acompanhamento, δ_i uma função indicadora de falha e x_i representa as covariáveis associadas a este indivíduo.

A taxa de falha no intervalo $[t, t+\Delta t)$ é definida como a probabilidade de que a falha ocorra neste intervalo, dado que não ocorreu antes de t , dividida pelo comprimento do intervalo. Desta forma, se $\Delta t \rightarrow 0$, $\lambda(t)$ representa a taxa de falha instantânea no tempo t , condicional à sobrevivência até o tempo t , e é expressa por:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

Modelos tanto paramétricos quanto semi-paramétricos estão disponíveis para modelar dados de sobrevivência (Colosimo e Giolo, 2006). Os modelos paramétricos Exponencial e Weibull são comumente usados por sua simplicidade e pela fácil interpretação de seus parâmetros. Na prática, entretanto, os modelos semi-paramétricos de riscos proporcionais (modelos de Cox, 1972) são amplamente usados pois, eles não impõe uma forma particular para a curva de sobrevivência.

2.1 Modelo de Riscos Proporcionais de Cox

No modelo semi-paramétrico de riscos proporcionais de Cox a função de risco assume a forma:

$$\lambda_i(t) = \lambda_0(t) \cdot \exp\{X_i^T \beta\} \quad (1)$$

em que, $\lambda_0(t)$, a função de taxa de falha básica e o vetor $X_i = (X_{1i}, X_{2i}, \dots, X_{pi})^T$ e β representam as p covariáveis medidas no i -ésimo indivíduo e seus correspondentes coeficientes de regressão.

O modelo de Cox, sem covariável dependente do tempo (1), assume que a razão das taxas de falha de dois indivíduos distintos na amostra é constante para todo o tempo de acompanhamento.

A estimação dos parâmetros deste modelo é baseada no **método da máxima verossimilhança parcial** (Collet, 2003; Cox, 1975):

$$L(\beta) = \prod_{i=1}^m \left(\frac{\exp\{X_i^T \beta\}}{\sum_{j \in R(t_i)} \exp\{X_j^T \beta\}} \right)^{\delta_i} \quad (2)$$

em que, $R(t_i)$ é o conjunto dos índices das observações sob risco no tempo t_i .

A função de verossimilhança parcial (2) assume que os tempos de sobrevivência são contínuos e, conseqüentemente, não pressupõe a possibilidade de empates nos tempos observados. Na prática, empates podem ocorrer nos tempos de falha ou de censura devido à escala de medida. Breslow (1972) e Peto (1972) propuseram uma modificação na função de verossimilhança parcial para acomodar as observações empatadas quando estas ocorrem.

Considere s_i o vetor formado pela soma das covariáveis para os indivíduos que falharam no mesmo tempo t_i ($i = 1, \dots, k$) e d_i o número de falhas neste tempo. A função de verossimilhança parcial proposta por estes autores tem a seguinte forma:

$$L(\beta) = \prod_{i=1}^k \frac{\exp\{s_i^T \beta\}}{\left(\sum_{j \in R(t_i)} \exp\{X_j^T \beta\} \right)^{d_i}} \quad (3)$$

Esta aproximação é adequada quando o número de empates em qualquer tempo não é grande. Quando não há empates a expressão (3) se reduz à (2).

2.2 Modelo de Cox com Covariáveis Dependentes do Tempo

Considere a possibilidade que entre as p covariáveis medidas em cada indivíduo, x_1, \dots, x_p exista alguma que foi medida n vezes, em cada indivíduo, ao longo do tempo. Assim temos $(T_i, \delta_i, X_i(t_j))$ o vetor que representa os dados de sobrevivência observados para o i -ésimo indivíduo no j -ésimo tempo; $X_i(t_j)$ o vetor de covariáveis do i -ésimo indivíduo no tempo j ($j = 1, \dots, n_i$).

Neste modelo a função de risco assume a forma:

$$\lambda_i(t) = \lambda_0(t_i) \cdot \exp\{X_i^T(t_i) \cdot \beta\} \quad (4)$$

Então, o modelo de Cox com variável dependente do tempo não supõe riscos proporcionais pois a razão das funções de risco de dois indivíduos passa a depender do tempo.

Assim, a função de verossimilhança parcial para o modelo de Cox com covariável dependente do tempo pode ser escrita como:

$$L(\beta) = \prod_{i=1}^m \left(\frac{\exp\{X_i^T(t_i) \cdot \beta\}}{\sum_{j \in R(t_i)} \exp\{X_j^T(t_i) \cdot \beta\}} \right)^{\delta_i} \quad (5)$$

Andersen e Gill (1982) mostraram que, sob certas condições de regularidade, os estimadores produzidos pela função de verossimilhança parcial (3 e 5) são consistentes e assintoticamente normais.

3. Análise de Dados Longitudinais

Esta seção introduz os conceitos associados à análise de dados longitudinais, a notação utilizada na literatura e apresenta os modelos marginal e de efeitos aleatórios com suas respectivas formas de estimação. Também conhecido como **medidas repetidas**, este tipo de estudo envolve bancos de dados em que a variável resposta, Y , é medida repetidamente em cada indivíduo da amostra, ao longo do tempo. Suponha um conjunto de m indivíduos avaliados n_i vezes ao longo do tempo. $Y_i = (Y_{i1}, \dots, Y_{in})$ é o vetor de respostas do i -ésimo indivíduo, então Y_{ij} representa a resposta do i -ésimo indivíduo no j -ésimo tempo.

Em análise de dados longitudinais assumimos que, entre indivíduos, as respostas são independentes, mas para o mesmo indivíduo elas são correlacionadas.

Diggle et al. (1996) mostraram que, ao modelar dados longitudinais pelo modelo linear normal as estimativas dos coeficientes de regressão, $\hat{\beta}'_s$, permanecem consistentes porém, estas são ineficientes uma vez que a variância dos coeficientes de regressão estimados, $Var(\hat{\beta})$, fica subestimada.

Uma forma de corrigir este problema é através da estimação dos componentes de variância. O modelo marginal e o modelo de efeitos aleatórios apresentam duas formas de estimação destes componentes e serão apresentados a seguir.

3.1 Modelo Marginal

No modelo marginal a média e a variância de Y são modeladas separadamente. Assumimos que a resposta, Y_i , é resultado de um processo Gaussiano n -variado com média $(X_i^T \beta)$ e matriz de covariância V_i . A variância de V_i incorpora a estrutura de correlação das medidas do mesmo indivíduo. Portanto, é necessário atribuir forma específica à V_i . As formas mais comuns de V_i são: **Correlação Uniforme** neste caso, assumimos que a correlação entre as medidas do mesmo indivíduo é constante; **Correlação Serial**, esta forma assume que a correlação diminui quando as medidas se afastam no tempo; e, **Não-estruturada** neste caso, é especificada uma covariância diferente para cada par de observações do indivíduo. A matriz de correlação dos resíduos padronizados, obtidos pelo ajuste do modelo linear normal para a resposta longitudinal versus o tempo desconsiderando as demais variáveis, é útil para determinar a estrutura de correlação das medidas do mesmo indivíduo (Diggle et al, 1996).

O modelo marginal pode ser escrito como:

$$Y_i = X_i^T \beta + \varepsilon_i \quad (6)$$

em que, $Y_i \sim N_n(X_i^T \beta, V_i)$, e $\varepsilon_i \sim N(0, V_i)$.

3.2 Modelo de Efeitos Aleatórios

Neste modelo a resposta é assumida ser uma função linear das variáveis explicativas com coeficientes de regressão que variam entre indivíduos, isto é,

este modelo permite que os coeficientes sejam fixos ou aleatórios. Os coeficientes aleatórios explicam a dependência entre as respostas do mesmo indivíduo, esta variabilidade reflete a heterogeneidade devido a fatores não mensurados ou não mensuráveis.

A premissa básica do modelo de efeitos aleatórios é que a heterogeneidade natural entre indivíduos é modelada em um subconjunto de coeficientes de regressão. O modelo de efeitos aleatórios assume a seguinte forma:

$$Y_i = X_i^T \beta + Z_i b_i + \varepsilon_i ,$$

em que, $X_i^T \beta$: representa o **componente fixo**; X_i são as covariáveis incluídas na análise e β 's seus respectivos coeficientes de regressão, $Z_i b_i$: representa o **componente aleatório**; b_i 's são os coeficientes aleatórios que variam para cada indivíduo; e, Z_i é, em geral, um subconjunto de X_i de covariáveis dependentes do tempo.

No modelo de efeitos aleatórios assumimos que as respostas do i -ésimo indivíduo têm distribuição Normal n -variada com média $(X_i^T \beta)$ e matriz de covariância $\Sigma = (Z_i V_i Z_i^T + \sigma^2 I)$ em que V_i é a matriz de covariâncias das medidas do mesmo indivíduo. Os coeficientes aleatórios, b_i , e os erros do modelo, ε_i , são independentes e têm distribuição Normal com média zero e variâncias V_i e $\sigma^2 I$, respectivamente.

3.3 Inferência para os Modelos Marginal e de Efeitos Aleatórios

Os coeficientes de regressão do modelo marginal e do modelo de efeitos aleatórios podem ser estimados através do método da máxima verossimilhança. Entretanto, este método produz um estimador viciado para a variância (Diggle et al, 1996). O método da máxima verossimilhança restrita corrige este problema e sua função é dada por:

$$L_R(\theta) = |X^T H^{-1} X|^{-1/2} \cdot L_V(\theta) \quad (7)$$

em que, $H = \sigma^2 I$ e $L_V(\theta)$ é a função de verossimilhança usual:

$$L_V(\theta) = \prod_{i=1}^m (2\pi)^{-n/2} |V_i|^{-1/2} (\sigma^2)^{-n/2} \cdot \exp \left\{ -\frac{1}{2\sigma^2} (Y_i - X_i \beta)^T V_i^{-1} (Y_i - X_i \beta) \right\} \quad (8)$$

O estimador de máxima verossimilhança restrita possui as mesmas propriedades que o estimador de máxima verossimilhança.

4. Modelagem Conjunta

4.1 O Modelo Conjunto por Wulfsohn e Tsiatis (1997)

A modelagem conjunta de dados longitudinais e de sobrevivência tem sido explorada recentemente por muitos autores. O modelo conjunto proposto por Wulfsohn e Tsiatis (1997) tem sido referência para muitos deles. Estes autores utilizaram um modelo de Cox para o processo de sobrevivência e a resposta longitudinal é tratada como covariável contínua cujo valor real é dado por um modelo linear de efeitos aleatórios.

Considere n indivíduos, seja Y_i uma covariável dependente do tempo, e cada indivíduo tem m_i medidas da covariável longitudinal nos tempos $(t_{ij}; t_{ij} \leq T_i)$. T_i é o tempo máximo de acompanhamento do paciente e δ_i uma função indicadora de falha.

Assumimos que Y_{ij} pode ser ajustada por um modelo linear de efeitos aleatórios.

$$Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \varepsilon_{ij}$$

em que, ε_{ij} tem distribuição Normal com média zero e variância σ_ε^2 , com $\text{cov}(\varepsilon_{ij}, \varepsilon_{ij'}) = 0, \forall j \neq j'$. O vetor β_i tem distribuição normal bivariada com vetor de médias $\beta = (\beta_{0i}, \beta_{1i})^T$ e matriz de covariância σ .

$\varepsilon_{ij} \perp \beta_{0i}$ e β_{1i} .

O risco de falha é então ajustado pelo modelo semi-paramétrico de Cox e o valor ajustado pelo modelo linear de efeitos aleatórios utilizado como valor real do marcador longitudinal, em lugar da medida observada Y_{ij} .

$$\begin{aligned} \lambda_i(t | \beta_i, Y_i, t_i) &= \lambda_0(t | \beta_i) \\ &= \lambda_0(t) \cdot \exp\{\gamma(\beta_{0i} + \beta_{1i} t)\} \end{aligned}$$

em que, $\lambda_0(t)$ é a função de risco baseline (não-paramétrica) que é avaliada a cada tempo de falha.

Neste modelo, assumimos que o tempo de censura é não informativo, sendo realizado independentemente do tempo de sobrevivência e da medida do marcador longitudinal.

4.2 O Modelo Conjunto por Henderson et. al. (2000)

Henderson et. al. (2000) propuseram um modelo generalizado para a análise conjunta que pode ser dividido em dois sub-modelos e um processo de ligação. Para as medidas longitudinais é assumido um sub-modelo linear com efeitos aleatórios, e para o processo de sobrevivência um sub-modelo semi-paramétrico de riscos proporcionais. O modelo conjunto assume que dado o processo latente normal bivariado, $W(t) = \{W_1(t), W_2(t)\}$, e as covariáveis avaliadas, os processos longitudinal e de sobrevivência são condicionalmente independentes.

Os sub-modelos podem ser escritos na seguinte forma:

(1) Sub-modelo para as medidas longitudinais:

$$Y_{ij} = \mu_i(t_{ij}) + W_{1i}(t_{ij}) + \varepsilon_{ij} \quad (9)$$

em que, $\mu_i(t_{ij})$ é a resposta média e, $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ é uma seqüência de erros mutuamente independentes. Assumimos que a resposta média pode ser descrita por um modelo linear:

$$\mu_i(t) = X_{1i}(t)' \beta_1$$

em que, os vetores $X_{1i}(t)$ e β_1 representam variáveis explicativas, que podem ser dependentes do tempo, e seus correspondentes coeficientes de regressão.

(2) Sub-modelo para o processo de sobrevivência no tempo t é dado pelo modelo semi-paramétrico multiplicativo:

$$\lambda_i(t) = H_i(t) \alpha_0(t) \exp\{X_{2i}(t)^T \beta_2 + W_{2i}(t)\} \quad (10)$$

em que, o componente não-paramétrico $\alpha_0(t)$ têm forma não especificada. $H_i(t)$ representa uma indicadora que é igual a 1 se o i -ésimo indivíduo está sob risco de evento no tempo t , e 0 caso contrário. Os vetores $X_{2i}(t)$ e β_2 podem, ou não, ter elementos em comum com $X_{1i}(t)$ e β_1 .

Para o processo de ligação latente assumimos que,

$$W_1(t) = U_1 + U_2 t \quad (11)$$

em que, (U_1, U_2) têm distribuição normal bivariada com médias zero, com respectivas variâncias σ_1^2 e σ_2^2 e coeficiente de correlação ρ . E, $W_2(t)$ tem a forma:

$$W_2(t) = \gamma_1 U_1 + \gamma_2 U_2 + \gamma_3 (U_1 + U_2 t) + U_3 \quad (12)$$

em que, $U_3 \sim N(0, \sigma_3^2)$ é independente de (U_1, U_2) . Neste modelo, os parâmetros γ_1 , γ_2 e γ_3 medem a associação induzida pelo intercepto, inclinação e valor de $W_1(t)$, respectivamente. U_3 modela a fragilidade ortogonal do processo de medidas.

5. Aplicação

5.1 Descrição do experimento

As conseqüências da infecção conjunta da malária com a esquistossomose têm sido relatadas por alguns autores. Em modelos experimentais, utilizando o *S. mansoni* (esquistossomose) em associação com espécies de *Plasmodium* (malária) de roedores, Lewin et al (1981) mostraram que as parasitoses mistas levaram a um aumento de parasitemia devida ao protozoário.

Com o objetivo de aprofundar os conhecimentos sobre o efeito da associação da infecção do protozoário da malária *P.berguei* com o helminto causador da esquistossomose *S.mansoni*, sobre a parasitemia por *P.berguei* e sobre a sobrevivência dos camundongos, foi conduzido, no Instituto de Ciências Biológicas (ICB-UFMG) em parceria com o Centro de Pesquisas René Rachou (CPqRR-FIOCRUZ), um experimento em que 80 camundongos foram alocados aleatoriamente em 4 grupos de tratamentos.

Estes camundongos foram acompanhados durante 25 dias. O tempo zero indica o tempo em que os camundongos foram infectados com *P. berguei*. A partir do 5º dia após a infecção o percentual de parasitas no sangue do camundongo torna-se grande o suficiente para ser detectado pelo teste de esfregaço sanguíneo (teste para determinar a parasitemia), este dia então foi usado como padrão para o início da contagem que foi realizada diariamente até o 30º dia, quando todos os camundongos que haviam sobrevivido foram sacrificados. Valores altos de parasitemia indicam alto grau de infecção do camundongo, enquanto que valores observados iguais a zero indicam a ocorrência de “cura” do camundongo. Os camundongos que morreram ao longo do período de acompanhamento (5º ao 30º dia após a infecção), tiveram o dia do óbito registrado.

Os 80 camundongos foram distribuídos aleatoriamente entre 4 grupos de 20 camundongos cada, no entanto, ocorreram mortes de camundongos antes do início da medida da parasitemia devido a fatores distintos do objetivo do estudo sendo, portanto, estes eliminados do banco de dados. O banco final dispõe de 64 camundongos assim distribuídos:

- Grupo 1:** 20 camundongos infectados por *P. berguei* que foram previamente imunizados à malária por infecções sucessivas seguidas de tratamento.
- Grupo 2:** 16 camundongos infectados cronicamente por *S. mansoni* e posteriormente infectados por *P. berguei* que foram previamente imunizados à malária por infecções sucessivas seguidas de tratamento.
- Grupo 3:** 15 camundongos infectados por *P. berguei*.
- Grupo 4:** 13 camundongos infectados cronicamente por *S. mansoni* e posteriormente infectados por *P. berguei*.

Indução à imunidade à *P. berguei*:

Os camundongos dos grupos 1 e 2 foram imunizados contra a malária por *P. berguei* através de sucessivas infecções experimentais seguidas de cura utilizando tratamento com cloroquina (100mg/kg, via oral). O processo de indução de imunidade à malária iniciou-se 18 semanas após a infecção por *S. mansoni*. Para comprovar o sucesso da imunização, um grupo de 10 camundongos foi posteriormente infectado com 1×10^6 hemácias parasitadas por *P. berguei* e acompanhado para a constatação da ausência de parasitemia.

5.2 Análise de dados

Com base no banco de dados deste experimento definimos $O_{ij} = (X1_i, X2_i, X3_{ij}, t_{ij}, T_i, \delta_i)$ os valores observados das medidas do *i*-ésimo camundongo no *j*-ésimo tempo. $X1_i$ e $X2_i$, representam as medidas das covariáveis indicadoras: coinfeção e imunização, respectivamente; $X3_{ij}$, representa o percentual médio de parasitemia do *i*-ésimo camundongo no *j*-ésimo dia após a infecção; t_{ij} o tempo das medidas longitudinais, T_i , representa o tempo máximo de acompanhamento do *i*-ésimo camundongo, $T_i \leq 30$; e, δ_i é o indicador de falha.

A Figura 1, apresentada no Anexo A, contém os gráficos de perfis dos dados associados a cada camundongo por grupo experimental. O grupo 1 (Figura 1 (a)) é composto por camundongos que foram imunizados e infectados

apenas por malária, correspondendo ao grupo de melhor prognóstico uma vez que apresenta os mais baixos valores de parasitemia e, a partir do 25º dia após a infecção até o tempo final do estudo todos os camundongos apresentaram parasitemia zero (indicativo de cura) sendo portanto, todos censurados no 30º dia após a infecção. No grupo 2 (Figura 1 (b)) estão alocados os camundongos que foram imunizados e coinfetados. Neste grupo temos a presença de 6 camundongos que mantiveram o percentual de parasitemia baixo até o final do experimento e foram censurados no 30º dia após a infecção, 8 que morreram com níveis de parasitemia diversos e 2 casos que tiveram óbito no 12º dia mas que apresentavam indicador de cura no dia anterior à morte. Os camundongos alocados nos grupos 3 e 4 (Figura 1 (c) e (d)) não foram imunizados e todos eles morreram antes do fim do estudo, sendo que podemos observar valores altos de parasitemia precedendo a morte. Observa-se que no grupo 4, em que os camundongos foram coinfetados, o tempo de sobrevivência de apenas 2 animais foi superior a 11 dias, enquanto que no grupo 3 (animais com infecção simples) o tempo de sobrevivência parece ter sido maior. Nos grupos experimentais 3 e 4 as medidas de parasitemia do 5º dia após a infecção foram perdidas.

5.2.1 Análise de Dados Longitudinais

Primeiramente, iremos focar a análise deste banco de dados sob o aspecto da estrutura de medidas repetidas. Então, a variável resposta a ser considerada, X_{3ij} , é o percentual médio de parasitemia do i -ésimo camundongo no j -ésimo dia após a infecção, com $i = 1, \dots, 64$ e $j = 5, \dots, n_i$; $5 \leq n_i \leq 30$.

A Figura 2 apresenta a curva suavizada (método Lowess – Cleveland, 1985) do percentual de parasitemia médio para cada um dos grupos experimentais. Observa-se que o nível de parasitemia dos camundongos que foram imunizados (grupos 1 e 2) é aparentemente inferior ao nível dos camundongos não imunizados (grupos 3 e 4). Além disso, observa-se que os camundongos dos grupos 3 e 4 apresentam valores de parasitemia crescentes ao longo do acompanhamento, enquanto que os camundongos alocados nos

grupos 1 e 2, apresentaram valores mais baixos de parasitemia em todo o período de acompanhamento. Os gráficos de perfis do percentual médio de parasitemia para cada um dos grupos experimentais foram traçados e estão apresentados no Anexo A.

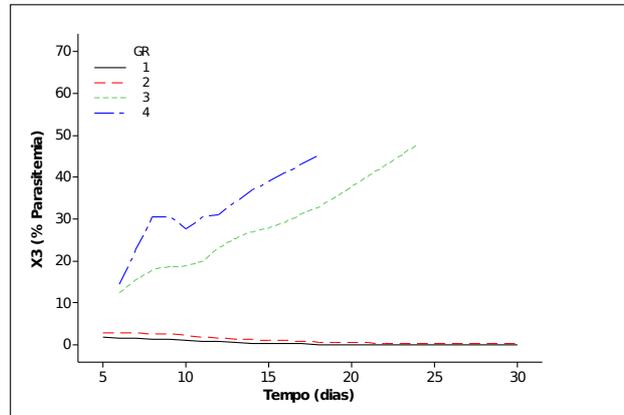


Figura 2 – Gráfico da curva de alisamento (Lowess) do % médio de parasitemia para os grupos experimentais.

Um modelo de regressão linear simples foi ajustado para a variável resposta: percentual de parasitemia (X3) versus o tempo após a infecção, desconsiderando-se o grupo experimental e o camundongo. Os resíduos padronizados foram calculados com o objetivo de avaliar a variabilidade da resposta no tempo (Liang, Diggle e Zeger, 1996). A partir destes resíduos, foi traçado um gráfico de perfis para cada grupo experimental. Estes gráficos estão apresentados no Anexo B e pode-se observar que, exceto no grupo 3 em que, a amplitude de variação dos resíduos parece constante, a variância dos resíduos parece aumentar com o tempo.

A matriz de correlação dos resíduos padronizados entre os tempos distintos estão apresentados no Anexo B, estas medidas foram calculadas no intuito de avaliar a estrutura de correlação das medidas no tempo, ou seja, se a correlação entre os tempos decresce quando as medidas se afastam sugerindo uma correlação serial ou se possuem outra forma como a correlação uniforme ou não-estruturada. Com base nos resultados encontrados observamos que a correlação tende a decrescer quando as medidas se afastam no tempo, logo a forma de correlação serial parece mais adequada para ajuste dos dados.

O modelo marginal e o modelo de efeitos aleatórios são as principais formas de modelagem de dados longitudinais e foram apresentadas na Seção 3. Uma série de modelos, utilizando estas técnicas, foram ajustados ao banco de dados e seguem descritos a seguir:

ML1 – Modelo Marginal com Correlação Uniforme: este modelo considera que a correlação entre as medidas do mesmo indivíduo é constante no tempo.

ML2 – Modelo Marginal com Correlação Serial: este modelo considera que a correlação entre as medidas do mesmo indivíduo decresce quando estas se afastam no tempo.

ML3 – Modelo Marginal com Correlação Não-estruturada: este modelo considera que não há estrutura na correlação das medidas do mesmo indivíduo em relação ao tempo.

ML4 – Modelo de Efeitos Aleatórios (efeito aleatório no intercepto): este modelo considera um intercepto distinto para cada camundongo, β_{0i} .

ML5 – Modelo de Efeitos Aleatórios (efeito aleatório no intercepto e no Tempo): este modelo permite que tanto o intercepto, quanto a inclinação com o tempo variem entre camundongos.

A Tabela 1 apresenta um resumo destes ajustes em que são apresentados os coeficientes de regressão estimados ($\hat{\beta}$'s) e os valores p associados ao teste de Wald. Nesta tabela são apresentadas duas medidas de qualidade do ajuste. São elas:

AIC: Akaike Information Criterion $AIC = -2 \left[l(\hat{\theta}) - \dim(\theta) \right]$

BIC: Bayesian Information Criterion

$$BIC = -2 \left[l(\hat{\theta}) - \frac{N \log(\dim(\theta))}{2} \right]$$

em que, θ representa o conjunto de parâmetros estimados em cada modelo.

Estas medidas são usadas para comparar modelos a partir da informação do quanto a função de verossimilhança foi capaz de absorver dos dados, penalizada pelo grau de complexidade do modelo. Valores menores destas medidas indicam melhores ajustes.

A partir dos resultados apresentados na Tabela 1, podemos confirmar que o modelo marginal com correlação serial (ML2) é o que melhor se ajustou aos dados uma vez que, apresentou os menores valores de AIC e BIC. Neste modelo todas as covariáveis foram significativas. A coinfeccção tem efeito de aumentar o grau de parasitemia em cerca de 3%, enquanto que a imunização reduz em média 22% o nível de parasitemia do camundongo. O efeito do tempo é de crescimento na parasitemia (em média 0,18% ao dia).

Tabela 1 – Ajuste de modelos para Dados Longitudinais

Modelo	Covariáveis no modelo	Estimativas	p-valor	AIC	BIC
ML1	Intercepto X1: coinfeccção X2: imunização T: tempo	$\beta_0 = 23,16$ $\beta_1 = 3,47$ $\beta_2 = -22,39$ $\beta_3 = 0,056$	0,0000 0,0021 0,0000 0,1181	6592,02	6621,14
ML2	Intercepto X1: coinfeccção X2: imunização T: tempo	$\beta_0 = 21,69$ $\beta_1 = 2,99$ $\beta_2 = -23,45$ $\beta_3 = 0,183$	0,0000 0,0209 0,0000 0,0143	6120,13	6149,24
ML4	Intercepto X1: coinfeccção X2: imunização T: tempo	$\beta_0 = 23,16$ $\beta_1 = 3,47$ $\beta_2 = -22,39$ $\beta_3 = 0,056$	0,0000 0,0030 0,0000 0,1180	6592,02	6621,14
ML5	Intercepto X1: coinfeccção X2: imunização T: tempo	$\beta_0 = -0,72$ $\beta_1 = 0,71$ $\beta_2 = -7,91$ $\beta_3 = 2,18$	0,7733 0,5573 0,0000 0,0000	6222,88	6261,69

* O ajuste do modelo ML3 (Modelo Marginal com correlação Não-estruturada) não é apresentado pois, devido à sua complexidade, este modelo não convergiu.

5.2.2 Análise de Sobrevivência

Agora, iremos focar a estrutura deste banco de dados sob o aspecto de análise de sobrevivência. Então, a variável resposta, T, é o tempo transcorrido entre a infecção e a morte camundongo, sendo que houve censura no tempo $T = 30$ dias, quando todos os camundongos que sobreviveram foram sacrificados.

Com o objetivo de se obter uma idéia do efeito das covariáveis sobre o comportamento dos dados foram calculadas, inicialmente, as estimativas para a função de sobrevivência pelo método de Kaplan-Meier para as covariáveis X1 e X2. A Figura 3 apresenta as curvas de sobrevivência estimadas para a covariável X1 (coinfecção). Comparando estas curvas pelo teste de Wilcoxon, observa-se que os camundongos infectados apenas com malária (infecção simples) possuem maior probabilidade de sobrevivência em relação aos que foram coinfectados ($p=0,0001$).

As curvas de sobrevivência estimadas pelo método de Kaplan-Meier para a covariável X2 (imunização) estão apresentadas na Figura 4. Observa-se que os camundongos que foram imunizados possuem maior probabilidade de sobrevivência em relação aos que não foram imunizados. O teste de Wilcoxon confirmou que há diferença significativa ($p<0,0001$) entre as curvas de sobrevivência dos camundongos segundo a imunização.

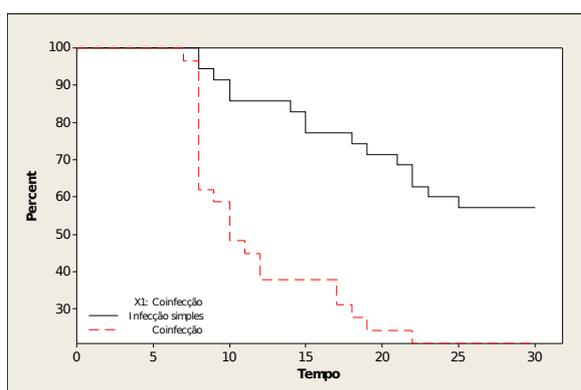


Figura 3 – Curvas de sobrevivência estimadas pelo método de Kaplan-Meier para a covariável X1: Co-Infecção.

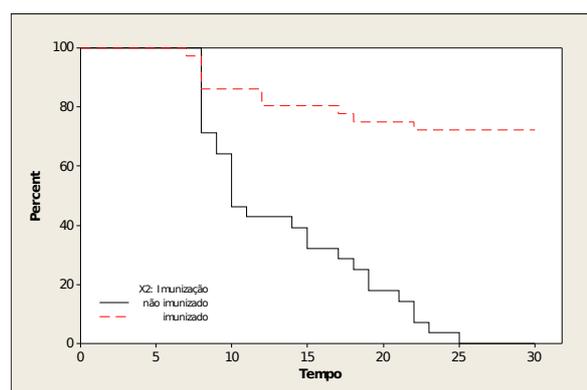


Figura 4 – Curvas de sobrevivência estimadas pelo método de Kaplan-Meier para o fator Imunização.

O modelo de riscos proporcionais de Cox (1972), apresentado na Seção 2, foi ajustado aos dados. A Tabela 2 apresenta um resumo dos resultados obtidos para o ajuste de modelos de Cox para as covariáveis X1 e X2.

Tabela 2 – Ajuste do Modelo de Cox

Modelo	Covariáveis no modelo	Estimativas	p-valor	exp{ β }	IC 95% (exp{ β })	
					LI	LS
1	X1: coinfeção	$\beta_1 = 1,14$	<0,001	3,13	1,61	6,07
2	X2: imunização	$\beta_2 = -1,99$	<0,001	0,137	0,0637	0,295
3	X1: coinfeção X2: imunização	$\beta_1 = 1,39$ $\beta_2 = -2,20$	<0,001 <0,001	3,997 0,111	1,9915 0,0498	8,022 0,249

A partir dos resultados apresentados na Tabela 2, observa-se que ambas as covariáveis Coinfeção (X1) e Imunização (X2) são significativas para explicar os tempos de sobrevivência dos camundongos. Os camundongos imunizados têm menor chance de morte e a presença de coinfeção aumenta a chance de mortalidade em cerca de 3 vezes.

6. Modelo de Cox com covariável dependente do tempo

O modelo de Cox com covariável dependente do tempo é um método bastante conhecido e utilizado na literatura. Portanto, consideramos este ajuste como a primeira aproximação para a modelagem conjunta de dados longitudinais e de sobrevivência. A Tabela 3 apresenta um resumo dos

resultados obtidos através do ajuste de deste modelo para as covariáveis X1, X2 e parasitemia, X3, que é dependente do tempo.

Tabela 3 – Ajuste do Modelo de Cox com covariável dependente do tempo

Covariáveis no modelo	Estimativas	p-valor	exp{ β }	IC 95% (exp{ β })	
				LI	LS
X1: coinfeção	$\beta_1 =$ 1,0520	0,0061	2,863	1,35	6,07
X2: imunização	$\beta_2 =$ -0,4814	0,38	0,618	0,21	1,79
X3: parasitemia	$\beta_3 =$ 0,0604	0,0000 062	1,062	1,04	1,09

Observa-se que a presença da covariável longitudinal no modelo acarreta a perda de significância da imunização (X2). Ou seja, ao acrescentar a parasitemia (X3) como covariável no modelo, esta absorve a informação do efeito de X2, tornando esta última não significativa. Como o objetivo principal do experimento é avaliar o efeito de tratamento sob a sobrevivência e sobre a parasitemia do camundongo, o modelo de Cox com covariável dependente do tempo torna-se ineficaz.

7. Referências Bibliográficas

- Andersen, P.K., Gill, R. (1982). Cox's Regression Model for Counting Processes: A Large Sample Study. *Annals of Statistics*, **10**, 1100-1200.
- Breslow, N. (1972). Contribuição à Discussão do artigo de D.R. Cox. *Journal of the Royal Statistical Society B*, **34**, 216-217.
- Brown, E.R., Ibrahim, J.G., DeGruttola, V. (2005). A Flexible B-Spline Model for Multiple Longitudinal Biomarkers and Survival. *Biometrics*, **61**, 64-73.
- Collet, D. (2003). *Modelling Survival Data in Medical Research*, 2ed., Chapman and Hall, London.
- Colosimo, E.A., Giolo, S.R. (2006). *Análise de Sobrevida Aplicada*. Edgard Blücher, São Paulo.
- Cox, D.R. (1972). Regression Models and Life Tables (with discussion). *Journal Royal Statistical Society, B*, **34**, 187-220.
- Cox, D.R. (1975). Partial Likelihood. *Biometrika*, **62**, 269-276.
- Diggle, P.J., Liang, K.Y., Zeger, S.L. (1996). *Analysis of Longitudinal Data*. Oxford University Press, New York.
- Guo, X., Carlin, P. (2004). Separate e Joint Modeling of Longitudinal and Event Time Data Using Standard Computer Packages. *The American Statistician*, **58**, 16-24.
- Henderson, R., Diggle, P., Dobson, A. (2000) Joint Modeling of Longitudinal Measurements and Event Time Data. *Biostatistics*, **1**, 465-480.
- Lewin, M., Last, C., Targett, G., Doenhoff, M.J. (1981). Factors affecting the acquisition of resistance against *Schistosoma mansoni* in the mouse. *Annals of Tropical Medicine and Parasitology*,

75, 79-86.

Peto, R. (1972). Contribuição à discussão do artigo de D.R. Cox. *Journal of the Royal Statistical Society*

B, **34**, 205-207.

Ratcliffe, S.J., Guo, W., Have, T.R.T. (2004). Joint Modeling of Longitudinal and Survival data via a

Common Frailty. *Biometrics*, **60**, 892-899.

Song, X., Davidian, M., Tsiatis, A.A. (2002). A Semiparametric Likelihood Approach to Joint

Modeling of Longitudinal and Time-to-Event Data. *Biometrics*, **58**, 742-753.

Tseng, Y.K., Hsieh, F., Wang, J.L. (2005). Joint Modeling of Accelerate Failure Time and

Longitudinal Data. *Biometrika*, **92**, 587-603.

Xu, J., Zeger, S.L. (2001). Joint Analysis of Longitudinal Data comprising Repeated measures and

Time to Events. *Royal Statistical Society*, **50**, 375-387.

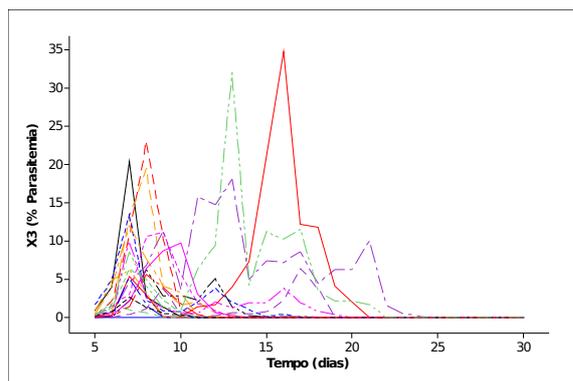
Wang, Y., Taylor, J.M.G. (2001). Jointly Modeling Longitudinal and Event Time Data with

Application to Acquired Immunodeficiency Syndrome. *The American Statistical association*, **96**, 895-905.

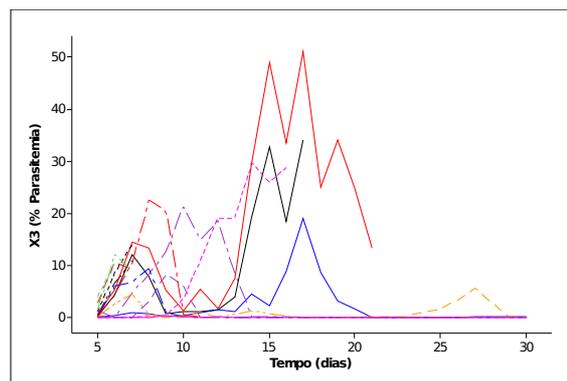
Zheng, Y., Heagerty, P.J. (2005). Partly Conditional Survival Models for Longitudinal Data.

Biometrics, **61**, 379-391.

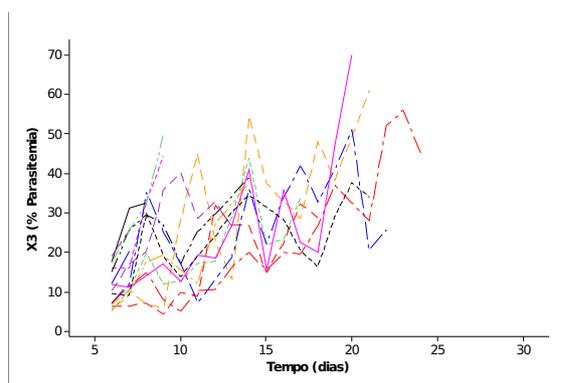
ANEXO A



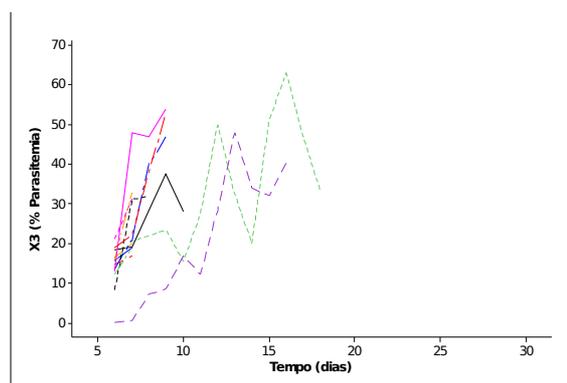
(a)



(b)



(c)



(d)

Figura 1 - Gráfico de Perfis do % médio de parasitemia para os camundongos por grupo experimental (a) grupo 1, (b) grupo 2, (c) grupo 3 e (d) grupo 4.

ANEXO B

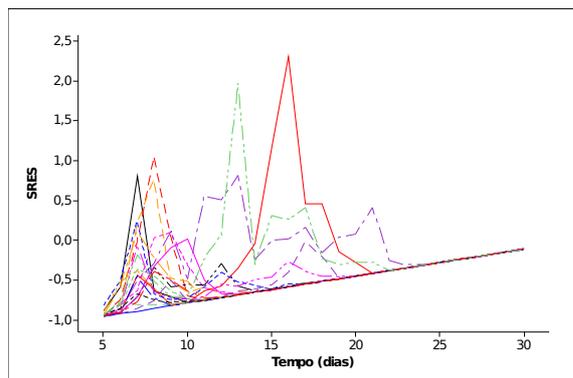


Gráfico de Perfis do resíduos padronizados para os camundongos do grupo 1

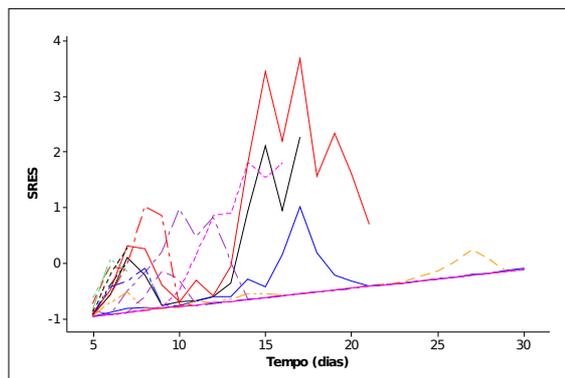


Gráfico de Perfis do resíduos padronizados para os camundongos do grupo 2

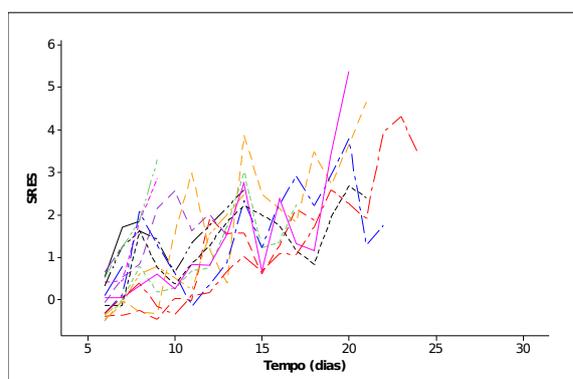


Gráfico de Perfis do resíduos padronizados para os camundongos do grupo 3

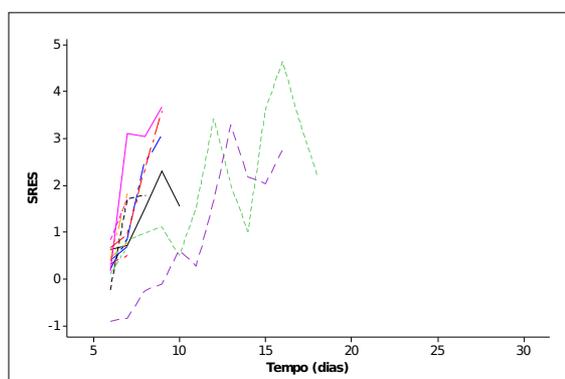


Gráfico de Perfis do resíduos padronizados para os camundongos do grupo 4

Matriz de correlação dos resíduos padronizados por tempo

	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	
6	0,74																									
7	0,45	0,80																								
8	0,04	0,81	0,80																							
9	0,12	0,85	0,75	0,91																						
10	0,35	0,61	0,41	0,56	0,79																					
11	0,36	0,64	0,43	0,44	0,61	0,84																				
12	0,30	0,67	0,41		0,72	0,77	0,82																			
13	0,1	0,56	0,34	0,53	0,64	0,63	0,68	0,87																		
14	0,05	0,62	0,37	0,56		0,75	0,78	0,71	0,76																	
15	0,09	0,58	0,42	0,45	0,46	0,53	0,66	0,68	0,66	0,80																
16	-0,0	0,65	0,38	0,48	0,58	0,59	0,68	0,78	0,76	0,80	0,92															
17	0,12	0,67	0,50	0,55	0,54	0,50	0,57	0,64	0,68	0,81	0,91	0,88														
18	0,09	0,70	0,44	0,49	0,54	0,70	0,78	0,71	0,66	0,92	0,85	0,86	0,89													
19	0,19	0,80	0,43	0,60	0,65		0,68	0,69		0,88		0,88														
20	0,20	0,82	0,39	0,59	0,69	0,67	0,71	0,74	0,71	0,95	0,71	0,84	0,76	0,85	0,97											
21	0,21	0,68		0,44	0,46		0,77	0,59	0,93	0,75	0,72	0,64	0,90	0,86	0,90	0,86	0,90									
22	0,10	0,61	0,28		0,45	0,32	0,42	0,43	0,46	0,78	0,61	0,56	0,64	0,84	0,91	0,84	0,91	0,84								
23	0,07	0,52		0,5	0,23	0,13	0,37	0,32	0,35	0,88	0,49	0,43	0,56	0,86	0,97	0,97	0,94	0,95								
24	-0,0	0,52	0,21	0,31	0,22	0,13	0,36		0,35	0,88	0,49	0,42		0,86	0,97	0,97		1								
25	0,78	0,3	0,6	0,8	0,6	0,1	0,5	0,32	0,4	0,4	0,3	0,9	0,56	0,1	0,7	0,8	0,94	0,9	0,33	0,99						
25	-0,0	-0,1	-0,2	-0,1	-0,1	-0,0	-0,0	-0,0	-0,0	-0,1	-0,0	-0,0	-0,0	-0,0	-0,0	-0,0	-0,0	-0,0	0,02	0,33	0,99					

