



UNIVERSIDADE FEDERAL DO PARANÁ
SETOR DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA
CURSO DE ESTATÍSTICA

ALINE RITER

MARIANA DAS MERCÊS CLARINDO

**MODELO DE COX DE EFEITOS MISTOS APLICADO A UM ESTUDO DE
HIPERTENSÃO EM FAMÍLIAS BRASILEIRAS**

CURITIBA

2011

ALINE RITER
MARIANA DAS MERCÊS CLARINDO

**MODELO DE COX DE EFEITOS MISTOS APLICADO A UM ESTUDO DE
HIPERTENSÃO EM FAMÍLIAS BRASILEIRAS**

Trabalho de Conclusão de Curso apresentado à disciplina Laboratório de Estatística, Curso de Estatística, do Setor de Ciências Exatas da Universidade Federal do Paraná, como requisito parcial à obtenção do título de Estatístico.

Orientadora: Prof^a Dr^a Suely Ruiz Giolo.

CURITIBA

2011

ALINE RITER

MARIANA DAS MERCÊS CLARINDO

**MODELO DE COX DE EFEITOS MISTOS APLICADO A UM ESTUDO DE
HIPERTENSÃO EM FAMÍLIAS BRASILEIRAS**

Este trabalho, que é requisito para a aprovação na disciplina Laboratório de Estatística do curso de Estatística da Universidade Federal do Paraná, foi avaliado pela seguinte banca examinadora:

COMISSÃO EXAMINADORA

Orientador

Prof.^a Dr.^a Suely Ruiz Giolo.

1º Membro

Prof.^a Dr.^a Nívea da Silva Matuda.

Curitiba, ____ de _____ de 2011.

AGRADECIMENTOS

Agradecimentos da Aline

Agradeço infinitamente aos meus pais Marcos e Tina por todo apoio durante todo esse trajeto, ao Saulo por toda compreensão e auxílio nos momentos difíceis, a minha amiga e companheira de TCC Mariana pela cumplicidade, aos meus amigos de sala e a professora Suely não só pela orientação neste semestre, mas por todo conhecimento e motivação transmitidos nestes anos de faculdade.

Agradecimentos da Mariana:

Agradeço à minha mãe Clarice, por todos os ensinamentos e apoio incondicional; à minha irmã Natália e a Tchenna pelos momentos tão importantes em família e por me ajudarem na adaptação em Curitiba; aos amigos de verdade conquistados em Curitiba e aos que ficaram em Tangará, em especial Paulo Neri e Nathalia Maria: presentes nos momentos mais felizes e mais difíceis.

Aos meus companheiros do Curso de Estatística – UFPR 2008, pela união e cooperativismo do dia-a-dia, em especial, à minha amiga e parceira deste trabalho Aline.

Aos professores que ensinaram e instigaram à minha curiosidade por aprender, desde o ensino fundamental até a Universidade. Em especial ao Prof.^o João Costalonga e Prof.^o José Luiz (Batata) grandes incentivadores da arte da Matemática, e à professora Suely, pela orientação deste trabalho, conselhos, conhecimentos, enfim, por ter transformado positivamente a minha maneira de enxergar a Estatística.

“Acreditar em algo e não viver é desonesto”

Mahatma Gandhi

RESUMO

Neste trabalho, foram analisados dados do Estudo Corações de Baependi, MG, Brasil, constituído por 1.673 indivíduos de 81 famílias. Este estudo foi realizado com o objetivo de identificar fatores de risco associados à doenças cardiovasculares, dentre elas a hipertensão arterial e a idade em que esta se manifesta, consideradas neste trabalho. Para que o parentesco existente entre os indivíduos de uma mesma família pudesse ser considerado nas análises, bem como a questão sobre a associação de fatores genéticos e ambientais com a idade em que a hipertensão arterial se manifesta pudesse ser respondida, foi utilizado o modelo de Cox de efeitos mistos. Com base nesse modelo, os componentes de variância associados aos efeitos poligênico e ambientais foram estimados, sendo que somente o efeito poligênico se mostrou significativo. De modo geral, os resultados apontaram que as covariáveis índice de massa corporal (IMC) e gênero, bem como fatores genéticos, estão associados à hipertensão e à idade em que esta se manifesta. Com tais conclusões, um importante espaço se abre para pesquisas futuras no sentido de identificar os genes que estariam associados a esta doença.

Palavras-chave: Modelo de Cox, efeitos mistos, efeitos genéticos, hipertensão.

LISTA DE FIGURAS

FIGURA 1: Árvore genealógica dos indivíduos que compõem a família 79. Círculos e quadrados representam indivíduos do gênero feminino e masculino, respectivamente, sendo estes preenchidos de preto quando houver diagnóstico positivo de hipertensão arterial. Códigos de identificação dos indivíduos e idade de diagnóstico dos hipertensos (ou idade na data da pesquisa dos normotensos) são mostrados na árvore.	6
FIGURA 2: Sobrevivência estimada por Kaplan-Meier. Os tempos representados por + mostram onde ocorreram censuras.....	14
FIGURA 3: Curvas de sobrevivência estimadas por Kaplan-Meier. Os tempos representados por + mostram onde ocorreram censuras em cada grupo.....	16
FIGURA 4: Resíduos padronizados de Schoenfeld do modelo de Cox com as covariáveis IMC e gênero.....	18
FIGURA 5: Representação gráfica do intervalo de confiança associado ao parâmetro σ_g^2 efeito poligênico do modelo 2.	19
FIGURA 6: Representação gráfica dos intervalos de confiança associados aos parâmetros σ_g^2 e σ_f^2 , efeitos poligênico e ambiental do modelo 3.	20
FIGURA 7: Estimativas dos efeitos genéticos g_{ij} associados aos indivíduos i ($i = 1, \dots, n_j$) de cada família j ($j = 1, \dots, 81$).	22
FIGURA 8: Boxplots das estimativas dos efeitos genéticos g_{ij} associados aos indivíduos i ($i = 1, \dots, n_j$) de cada família j ($j = 1, \dots, 81$).	22
FIGURA 9: Árvore genealógica dos indivíduos da família 48, com suas respectivas idades de diagnóstico de hipertensão (ou na data de pesquisa para normotensos) e risco de hipertensão	23
FIGURA 10: Árvore genealógica dos indivíduos da família 45, com suas respectivas idades e risco de hipertensão.....	23
FIGURA 11: Representação das funções de risco acumulada estimadas para os indivíduos das famílias 48 (em vermelho) e 45 (em azul) e para um indivíduo com efeito genético $g_{ij} = 0$	24

LISTA DE TABELAS

TABELA 1: Covariáveis eleitas para as análises estatísticas.....	7
TABELA 2: Frequências e prevalência de hipertensão de acordo com as covariáveis dicotomizadas.	15
TABELA 3: Resultados do teste logrank utilizado para comparar as curvas de sobrevivência obtidas para as categorias de cada covariável considerada.	17
TABELA 4: Estimativas obtidas para os parâmetros nos modelos de Cox ajustados	17
TABELA 5: Testes de proporcionalidade dos riscos no modelo de Cox	18
TABELA 6: Estimativas obtidas para os parâmetros nos modelos de Cox de Efeitos Mistos	19
TABELA 7: Resultados do ajuste do modelo de Cox de Efeitos Mistos	20

SUMÁRIO

1. INTRODUÇÃO	1
2. MATERIAL E MÉTODOS	4
2.1 MATERIAL	4
2.2 MÉTODOS	8
2.2.1 Estimador de Kaplan-Meier e Teste <i>Logrank</i>	8
2.2.2 Modelo de Regressão de Cox	9
2.2.3 Modelo de Cox de Efeitos Mistos	11
3. RESULTADOS	14
3.1 ANÁLISE DESCRITIVA	14
3.2 AJUSTE DE MODELOS	17
4. CONCLUSÃO E DISCUSSÃO	25
REFERÊNCIAS	27
ANEXO	29
ANEXO 1 – Obtenção da Figura 5 – Intervalo de confiança	29

1. INTRODUÇÃO

Popularmente conhecida como pressão alta, a hipertensão arterial é uma doença cardiovascular caracterizada pelo aumento da pressão do sangue contra a parede das artérias, oferecendo resistência à passagem do sangue que é bombeado pelo coração para irrigar os órgãos ou movimentar-se. Este bombeamento é composto por dois movimentos: o primeiro de contração, força máxima para expulsar o sangue e o segundo de relaxamento do coração entre as contrações cardíacas, de força mínima, denominados sístole e diástole, respectivamente. Equipamentos como estetoscópio e esfigmomanômetro, com escala em milímetros de mercúrio (mmHG), são utilizados por profissionais de saúde treinados para aferir a pressão arterial. A linha demarcatória que define se um indivíduo possui hipertensão arterial considera valores acima de 140mmHG para a pressão arterial sistólica e/ou acima de 90mmHG para a diastólica, 140/90mmHG, sendo que este diagnóstico deverá sempre ser validado por medidas repetidas em pelo menos três ocasiões (AZEVUN & GUIMARÃES, 2010).

A hipertensão é considerada uma doença silenciosa, pois seus sintomas como dor de cabeça, tontura, cansaço, entre outros, quando observados, são comuns e se confundem com sintomas de outras doenças. Porém se não tratada, é considerada um dos principais fatores de risco de mortalidade por doenças cardiovasculares, pois afeta órgãos vitais podendo ocasionar graves problemas de saúde como derrames cerebrais, doenças do coração, insuficiência renal, paralisção dos rins e alterações na visão que podem levar à cegueira, tornando-se assim um dos grandes problemas de saúde pública no Brasil e no mundo (CARVALHO, 2006 *apud* AZEVUN & GUIMARÃES, 2010). Em 2001, cerca de 7,6 milhões de mortes no mundo foram atribuídas à elevação da pressão arterial, 54% por acidente vascular encefálico e 47% por doença isquêmica do coração. No Brasil, doenças cardiovasculares ainda são responsáveis pela alta frequência em internações e também a principal causa de morte, ocasionando custos médicos e socioeconômicos elevados. Em novembro de 2009 ocorreram 91.970 internações por doenças cardiovasculares, resultando em um custo de R\$

165.461.644,33 (DATASUS, 2009 *apud* AZEVUN & GUIMARÃES, 2010). Para minimizar estes problemas que podem ocorrer pela elevação da pressão arterial, ações de orientação, alertando a população sobre os fatores de risco, prevenção e como diagnosticar a doença são realizadas continuamente por iniciativas públicas e privadas, podendo citar atualmente, a campanha “sou 12 por 8”, no Brasil.

Em todo o mundo, diversos grupos de pesquisa multidisciplinares, agregando profissionais de saúde, alunos, professores e pesquisadores de universidades foram constituídos com o objetivo comum de estudar e identificar os fatores de risco associados a doenças coronarianas. Os fatores de risco comumente associados à hipertensão arterial são relacionados aos maus hábitos alimentares (como o sal em excesso), ao sedentarismo, a obesidade, consumo excessivo de álcool e também a fatores genéticos (AZEUN & GUIMARÃES, 2010).

A descoberta de que fatores genéticos podem estar associados com doenças cardiovasculares têm motivado diversas pesquisas em genética com humanos, onde um dos objetivos principais é entender a arquitetura genética de regulação de fenótipos ou desordens complexas. Doenças identificadas como o produto da ação de múltiplos genes e fatores ambientais, bem como suas interações, são denominadas doenças ou desordens complexas. Para determinar os genes associados a estas doenças, dentre elas a hipertensão arterial, delineamentos em que a coleta de dados é realizada em famílias têm demonstrado superioridade. Tal superioridade se deve a resultados mais consistentes e com pouca ocorrência de falsos positivos, quando comparados com estudos que não identificam as relações de parentesco entre os pesquisados (THOMAS *et al.*, 2005; KULLO & DING, 2007). Como exemplos de estudos que consideram a estrutura das famílias para análise destes fatores, podem ser citados o *San Antonio Family Heart Study*, que utilizou uma amostra de famílias americanas mexicanas (MAHANEY *et al.*, 1995), e o *GENOA Study*, que envolve famílias americanas africanas e brancas não hispânicas (KULLO *et al.*, 2006).

No Brasil, o primeiro estudo, e único até o momento, a registrar dados de famílias brasileiras com o mesmo objetivo dos dois estudos citados anteriormente é o “Estudo Corações de Baependi, MG” (OLIVEIRA *et al.*,

2008). Este estudo apresenta, contudo, uma característica importante que o diferencia dos demais, a de envolver uma população miscigenada como é a brasileira, resultante de mais de cinco séculos de intercruzamentos étnicos, sendo uma das populações mais heterogêneas do mundo (PIMENTA *et al.*, 2006; PENA, 2000).

Neste trabalho, os dados desse estudo são analisados com o objetivo de identificar se fatores genéticos e ou ambientais estão associados com a idade em que a hipertensão arterial se manifesta. Como a idade exata em que a hipertensão se manifesta é desconhecida, uma vez que os sintomas aparecem no decorrer do tempo, as associações de interesse são avaliadas em relação à *idade de diagnóstico* de hipertensão. Para esse fim, é utilizado um conjunto de técnicas estatísticas de Análise de Sobrevivência, em particular o Modelo de Cox de Efeitos Mistos. Este modelo permite incorporar em sua estrutura não somente a correlação genética existente entre os indivíduos de uma mesma família, mas também as informações dos indivíduos que não desenvolveram hipertensão arterial até o término do estudo, usualmente denominadas censuras por serem informações parciais da resposta. Tais metodologias estatísticas, bem como os dados do Estudo Corações de Baependi, MG, são descritos em mais detalhes a seguir.

2. MATERIAL E MÉTODOS

2.1 MATERIAL

O banco de dados utilizado neste trabalho é advindo do Estudo Corações de Baependi, MG, que foi conduzido por pesquisadores do Laboratório de Genética e Cardiologia Molecular (LGCM) do Instituto do Coração (InCor/USP), com a finalidade de identificar fatores genéticos e genes associados a doenças cardiovasculares. As informações contidas nesse banco foram coletadas no período de dezembro de 2005 a janeiro de 2006 utilizando um plano amostral que consistia em selecionar aleatoriamente onze setores censitários dentre os trinta no qual a cidade de Baependi é subdividida (IBGE-2000 *apud* PEREIRA *et al.*, 2006). Destes onze setores, as residências foram amostradas por amostragem sistemática com fator de repetição vinte, o ponto inicial de cada setor foi selecionado ao acaso. As informações foram coletadas inicialmente de qualquer indivíduo morador da residência com idade igual ou superior a 18 anos que aceitasse participar do estudo. Definido o respondente, foi aplicado o questionário de coleta de dados familiares, identificando todos os parentes de primeiro, segundo e terceiro graus e respectivos parentes por união conjugal. A partir destas informações foram construídas árvores genealógicas das famílias amostradas, sendo todos os membros com idade igual ou superior a 18 anos convidados a participar do estudo. Para atrair estes participantes houve divulgação através de autoridades municipais e religiosas, jornais, médicos, contatos telefônicos, entre outros. Com a aplicação do plano amostral foram selecionadas 119 famílias totalizando 1.712 indivíduos. Destas 119 famílias, 38 eram constituídas por indivíduos que não estabeleciam entre si relações consanguíneas. Sendo assim, foram considerados para análise os dados de 81 famílias, totalizando 1.673 indivíduos com idade variando entre 18 e 100 anos. O protocolo de pesquisa do estudo foi aprovado pelo comitê de ética do Hospital das Clínicas, Universidade de São Paulo, Brasil, tendo cada participante assinado um termo de consentimento livre e esclarecido para autorizar sua participação no estudo.

Para cada indivíduo foi aplicado um questionário contendo informações demográficas, histórico médico e fatores de riscos ambientais. Para a realização dos exames físicos selecionou-se um posto de saúde de fácil acesso aos participantes. As medidas antropométricas (altura, peso, entre outras) foram coletadas seguindo procedimentos padronizados. Entre os exames foram realizados também o de medida da pressão arterial sanguínea para a identificação da hipertensão arterial e exames de sangue para identificação de diabetes e colesterol alto. O exame da medida da pressão arterial sanguínea foi realizado três vezes para cada indivíduo com intervalo mínimo de 3 minutos entre uma leitura e outra. Foram considerados hipertensos os indivíduos que apresentaram a média das três leituras da pressão arterial sanguínea sistólica e ou diastólica maior ou igual a 140/90mmHG. A prevalência de hipertensão no estudo foi de 35,6%, sendo este resultado similar a prevalência encontrada na população brasileira (PEREIRA *et al.*, 2006).

Como o objetivo deste trabalho é avaliar se fatores genéticos e ou ambientais estão associados com a idade em que a hipertensão arterial se manifesta, foi considerado como variável resposta a idade em que a hipertensão foi diagnosticada, uma vez que a idade exata em que esta se manifesta não é conhecida exatamente. Desse modo, para aqueles indivíduos em que a hipertensão arterial havia sido diagnosticada em data anterior à do estudo, foi considerada a idade de diagnóstico informada por eles. Já para os indivíduos que obtiveram o diagnóstico da doença durante a realização dos exames, a idade considerada foi aquela registrada na data da coleta da informação. Ainda, as respectivas idades dos indivíduos normotensos, isto é sem diagnóstico de hipertensão arterial até a data do estudo, foram consideradas como censuras.

Para ilustrar a estrutura dos dados, a Figura 1 mostra a árvore genealógica (ou heredograma) da família 79 composta de sete indivíduos.

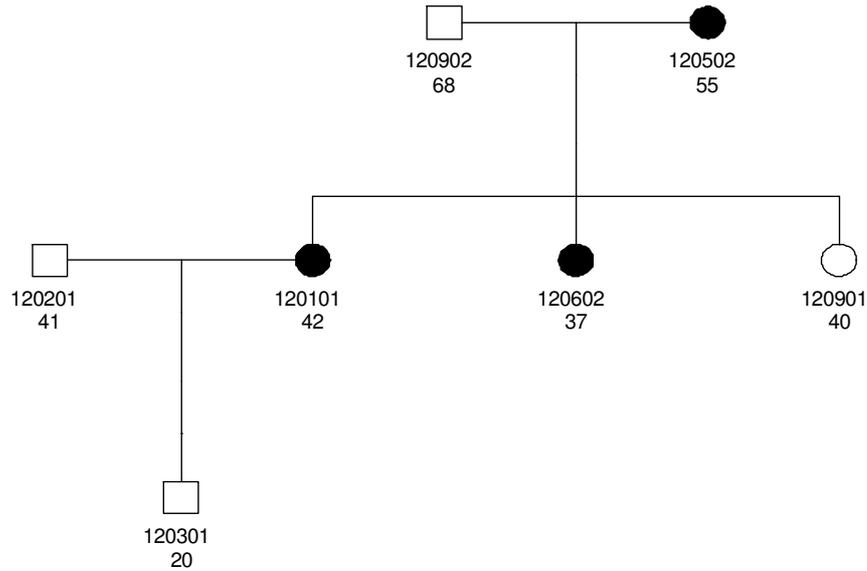


FIGURA 1: Árvore genealógica dos indivíduos que compõem a família 79. Círculos e quadrados representam indivíduos do gênero feminino e masculino, respectivamente, sendo estes preenchidos de preto quando houver diagnóstico positivo de hipertensão arterial. Códigos de identificação dos indivíduos e idade de diagnóstico dos hipertensos (ou idade na data da pesquisa dos normotensos) são mostrados na árvore.

Com essa estrutura de dados é possível mensurar a relação genética entre os indivíduos construindo a matriz de correlação genética $2\phi_j$ a seguir, sendo j ($j=1, \dots, k$) o indicador de família.

$$2\phi_j = \begin{matrix} i'_1 \\ i'_2 \\ \vdots \\ i'_n \end{matrix} \begin{bmatrix} i_1 & i_2 & \dots & i_{nj} \\ \varphi_{11} & \varphi_{12} & \dots & \varphi_{1n} \\ \varphi_{21} & \varphi_{22} & \dots & \varphi_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{n1} & \varphi_{n2} & \dots & \varphi_{nj} \end{bmatrix}_{n_j \times n_j}$$

em que:

$$\varphi_{ii'} = \begin{cases} 0 & \text{para } i \text{ não relacionado geneticamente com } i' \\ 1 & \text{para } i = i' \\ \left(\frac{1}{2}\right)^r & \text{para } i \neq i' \text{ e geneticamente relacionados,} \end{cases}$$

sendo r o grau de parentesco ($r = 1, 2, 3, \dots$). Esta matriz pode ser visualizada a seguir considerando a família 79.

$$2\phi_{79} = \begin{matrix} & \mathbf{120101} & \mathbf{120101} & \mathbf{120301} & \mathbf{120502} & \mathbf{120602} & \mathbf{120901} & \mathbf{120902} \\ \mathbf{120101} & 1 & 0 & 0,5 & 0,5 & 0,5 & 0,5 & 0,5 \\ \mathbf{120201} & 0 & 1 & 0,5 & 0 & 0 & 0 & 0 \\ \mathbf{120301} & 0,5 & 0,5 & 1 & 0,25 & 0,25 & 0,25 & 0,25 \\ \mathbf{120502} & 0,5 & 0 & 0,25 & 1 & 0,25 & 0,5 & 0 \\ \mathbf{120602} & 0,5 & 0 & 0,25 & 0,5 & 1 & 0,5 & 0,5 \\ \mathbf{120901} & 0,5 & 0 & 0,25 & 0,5 & 0,5 & 1 & 0,5 \\ \mathbf{120902} & 0,5 & 0 & 0,25 & 0 & 0,5 & 0,5 & 1 \end{matrix}$$

Exemplificando as relações de parentesco a partir da matriz de correlação genética apresentada, tem-se para os indivíduos 120502 e 120101, que são mãe e filha e, portanto, parentes de 1º grau, que a correlação entre eles é igual a 0,5. Já para os indivíduos 120602 e 120301, que são tia e sobrinho (parentes de 2º grau), a correlação é igual a 0,25.

A partir de matrizes $2\phi_j$ similares pode-se apresentar as correlações genéticas de todas as famílias utilizando-se uma matriz bloco diagonal, sendo esta quadrada com dimensão igual ao número de participantes do estudo, 1.673, mais os indivíduos que não estão no estudo, porém sua identificação faz-se necessária para a construção das relações de parentesco dentro de cada família, totalizando 2.260 indivíduos. A diagonal principal desta matriz é composta pelas matrizes $2\phi_j$ de correlações das famílias ($j=1,\dots,81$). Os elementos fora desta diagonal são nulos, dado que não existe correlação entre as famílias (DUCROCQ & CASSELA, 1996).

$$2\phi = \begin{bmatrix} 2\phi_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & 2\phi_{81} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} \varphi_{11} & \cdots & \varphi_{16} \\ \vdots & \ddots & \vdots \\ \varphi_{61} & \cdots & \varphi_{66} \end{bmatrix}_{6 \times 6} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \begin{bmatrix} \varphi_{51} & \cdots & \varphi_{55} \end{bmatrix}_{5 \times 5} \end{bmatrix}_{2260 \times 2260}$$

Dentre as covariáveis disponíveis no banco de dados, e considerando o conhecimento dos pesquisadores do LGCM do InCor/USP quanto àquelas que possivelmente apresentam relação com a idade de desenvolvimento da hipertensão arterial, foram eleitas para as análises estatísticas as quatro apresentadas na Tabela 1 a seguir.

TABELA 1: Covariáveis eleitas para as análises estatísticas.

Variável	Domínio
Gênero	Masculino e Feminino
Circunferência Abdominal	Centímetros
Índice de massa corporal (IMC)	Peso/Altura ²
Sedentarismo	Sim e Não

A seguir são apresentadas as metodologias estatísticas utilizadas para a análise dos dados aqui descritos.

2.2 MÉTODOS

Técnicas de análise de sobrevivência podem ser utilizadas quando o objetivo é analisar o tempo até a ocorrência de um evento de interesse, denominado tempo de falha, que pode ser o tempo até o óbito, a cura ou a recidiva de uma doença, entre outros. A principal característica dos dados de sobrevivência é a presença de censuras, que são informações parciais da resposta. Uma censura pode ocorrer por diversos motivos, dentre eles, a perda de acompanhamento do indivíduo e a não ocorrência do evento de interesse.

Assim como em outras situações, é importante realizar inicialmente uma análise exploratória dos dados de sobrevivência. No contexto de análise de sobrevivência, tal análise é usualmente feita utilizando-se o estimador de Kaplan-Meier, o teste *logrank* e medidas descritivas tais como o tempo médio e mediano. Modelos de regressão, em um contexto paramétrico ou semi-paramétrico, são também ferramentas úteis para a análise de dados de sobrevivência (COLOSIMO & GIOLO, 2006).

A seguir, são descritas, brevemente, as metodologias estatísticas de análise de sobrevivência utilizadas neste trabalho. Ênfase é dada ao modelo de regressão semi-paramétrico de Cox, bem como a uma extensão deste para dados correlacionados, que atende à estrutura dos dados do Estudo Corações de Baependi, MG.

2.2.1 Estimador de Kaplan-Meier e Teste *Logrank*

O estimador de Kaplan-Meier (KAPLAN & MEIER, 1958) é utilizado para estimar a função de sobrevivência, denotada por:

$$S(t) = P(T \geq t),$$

e que representa a probabilidade de uma observação não falhar (sobreviver) ao tempo t . Tal estimador considera tanto as falhas quanto as censuras na sua construção e é um dos estimadores mais utilizados na literatura clínica. A expressão do estimador de Kaplan-Meier é definida como:

$$\hat{S}(t) = \prod_{j:t_j < t} \left(1 - \frac{d_j}{n_j}\right)$$

com $t_1 < t_2 < \dots < t_k$ os k tempos distintos e ordenados de falha, d_j o número de falhas em t_j , $j= 1, 2, \dots, k$, e n_j o número de indivíduos sobre risco em t_j .

Dentre as propriedades deste estimador, as principais são: é não viciado para amostras grandes, fracamente consistente, converge assintoticamente para um processo gaussiano e é o estimador de máxima verossimilhança de $S(t)$.

As estimativas $S(t)$, obtidas a partir do estimador de Kaplan-Meier, podem ser representadas graficamente, de modo a se ter uma curva de sobrevivência associada a cada uma das categorias das covariáveis de interesse. Para testar se as curvas $S(t)$ associadas à estas categorias diferem entre si, pode-se utilizar o teste *logrank* (MANTEL, 1966), particularmente quando a razão das funções de risco dos grupos a serem comparados for aproximadamente constante (propriedade de riscos proporcionais).

A estatística deste teste é baseada na diferença entre o número observado de falhas em cada grupo e uma quantidade que, para muitos propósitos, pode ser pensada como o correspondente número esperado de falhas sob a hipótese nula, H_0 : igualdade entre as curvas de sobrevivência, isto é, $H_0: S_1(t) = S_2(t)$ para todo t , sendo $S_1(t)$ e $S_2(t)$ as curvas de sobrevivência associadas às duas categorias sendo comparadas. Para grandes amostras, esta estatística de teste segue uma distribuição qui-quadrado com 1 grau de liberdade (COLOSIMO & GIOLO, 2006).

2.2.2 Modelo de Regressão de Cox

O modelo de regressão de Cox (COX, 1972) é comumente utilizado em estudos clínicos e abriu uma nova fase na modelagem de dados. A principal razão para esta popularidade é a presença do componente não-paramétrico tornando este modelo mais flexível quando comparado com os modelos paramétricos, que consideram uma distribuição de probabilidade associada à variável resposta, como a exponencial, a Weibull e a Log-Normal.

A expressão geral do modelo de Cox é dada por:

$$\lambda(t) = \lambda_0(t)g(x'\beta),$$

com $\mathbf{x} = (x_1, \dots, x_p)'$ denotando um vetor de p covariáveis e g uma função não negativa tal que $g(0) = 1$, usualmente expressa por:

$$g(\mathbf{x}'\boldsymbol{\beta}) = \exp\{\mathbf{x}'\boldsymbol{\beta}\} = \exp\{\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p\},$$

em que $\boldsymbol{\beta}$ é um vetor de parâmetros associados às covariáveis. Desta forma, o modelo de Cox é composto por um componente não paramétrico, $\lambda_0(t)$, que é denominado função de risco de base ou basal e pelo componente paramétrico, $g(\mathbf{x}'\boldsymbol{\beta})$. Este modelo é também denominado modelo de riscos proporcionais, pois a razão das taxas de falha de dois indivíduos diferentes é constante, ou seja, a razão entre os riscos dos indivíduos i e j , a seguir,

$$\frac{\lambda_i(t)}{\lambda_j(t)} = \frac{\lambda_0 \exp\{x_i'\boldsymbol{\beta}\}}{\lambda_0 \exp\{x_j'\boldsymbol{\beta}\}} = \exp\{x_i'\boldsymbol{\beta} - x_j'\boldsymbol{\beta}\},$$

não depende do tempo. Por esta razão, a suposição básica deste modelo é de que as taxas de falha são proporcionais.

O ajuste do modelo de Cox é caracterizado a partir da obtenção dos coeficientes β 's e do termo não-paramétrico $\lambda_0(t)$, que podem ser estimados pelo método da máxima verossimilhança parcial (COX, 1975) e pelo estimador de Nelson-Aalen-Breslow (BRESLOW, 1972), respectivamente. Após o ajuste do modelo, as estimativas por ponto e por intervalo de cada coeficiente podem ser interpretadas, sendo que para a obtenção dos intervalos de confiança é necessário o cálculo do erro-padrão do vetor $\hat{\boldsymbol{\beta}}$. Para exemplificar, seja x_1 a covariável gênero, categorizada tal que 0 se feminino e 1 se masculino, e a variável resposta o tempo até a morte do paciente por câncer de laringe; a interpretação para a estimativa pontual $\exp(\hat{\beta}_1) = 2$, e seu correspondente intervalo de confiança (1,5; 2,8), seria que o risco de morte para mulheres é duas vezes o risco dos pacientes homens, com um intervalo de 95% de confiança de 1,5 a 2,8.

Para avaliar a adequação do modelo, deve-se verificar essencialmente o pressuposto de riscos proporcionais e também a qualidade global do ajuste. Para estas finalidades podem ser utilizados testes e métodos gráficos baseados nos resíduos de Schoenfeld (COLOSIMO & GIOLO, 2006 *apud* SCHOENFELD, 1982), bem como gráficos da função de risco acumulada construídos a partir das estimativas de risco.

2.2.3 Modelo de Cox de Efeitos Mistos

O modelo de regressão de Cox descrito previamente é útil para modelar dados em que os indivíduos são independentes. Desse modo, não é apropriado para analisar dados em que os indivíduos apresentem alguma estrutura de correlação, como é o caso dos dados do estudo tratado neste trabalho, que foi planejado e delineado respeitando a estrutura familiar dos indivíduos.

Para incorporar a estrutura de correlação mencionada, foi proposta uma extensão do modelo de Cox para dados correlacionados. Este modelo é denominado modelo de Cox de Efeitos Mistos (PANKRATZ et al., 2005; GIOLO et al., 2009) e é expresso por:

$$\lambda_{ij}(t) = \lambda_0(t) \exp\{\mathbf{x}'\boldsymbol{\beta} + \mathbf{g}_j\},$$

em que $\lambda_0(t)$ é a função de risco basal, $\boldsymbol{\beta}$ o vetor de coeficientes associados às covariáveis \mathbf{x} e $\mathbf{g}_j = (g_{1j}, g_{2j}, \dots, g_{n_{ij}})$ um vetor de efeitos aleatórios associado aos indivíduos i na família j ($i = 1, \dots, n_j$, $j = 1, \dots, k$). Usualmente é assumido que o vetor de efeitos aleatórios $\mathbf{g} = (\mathbf{g}_1, \dots, \mathbf{g}_k)$ segue uma distribuição normal multivariada tal que $\mathbf{g} \sim N_M(\mathbf{0}, \boldsymbol{\Sigma})$, sendo $\boldsymbol{\Sigma}$ a matriz de variância-covariância.

A matriz $\boldsymbol{\Sigma}$, no contexto dos dados do Estudo Corações de Baependi, MG, pode ser escrita como $\boldsymbol{\Sigma} = 2\phi\sigma_g^2$, em que 2ϕ denota a matriz de correlações genética (parentesco), descrita na Seção 3.1, e σ_g^2 o componente de variância poligênico associado ao efeito genético compartilhado pelos indivíduos. Outra sugestão para a matriz $\boldsymbol{\Sigma}$ é expressá-la por $\boldsymbol{\Sigma} = 2\phi\sigma_g^2 + \Omega\sigma_f^2$, em que a matriz Ω é denominada matriz de ambiente familiar compartilhado e σ_f^2 o componente de variância associado ao efeito ambiental compartilhado pelos indivíduos. A matriz bloco diagonal Ω considera em sua estrutura valores unitários dentro de cada bloco formado por indivíduos da mesma família. A justificativa para a construção desta matriz unitária e sua utilização no processo de estimação dos parâmetros para o modelo de Cox de efeitos mistos é advinda da suposição de que indivíduos de uma mesma família compartilham

costumes similares quanto à hábitos alimentares, prática de esportes, dentre outros.

Decorrente da estrutura do modelo proposto, o processo de estimação, tanto dos parâmetros β quanto dos componentes do vetor g , se torna extremamente complexo, pois a função de verossimilhança expressa por $L = \int PL(\beta, g) \frac{1}{\sqrt{2\pi|\Sigma|}} \exp\left[-\frac{1}{2}g'\Sigma^{-1}g\right] dg$, em que $PL = \text{parcial likelihood}$ (função de verossimilhança parcial), apresenta uma integral multidimensional intratável por métodos tradicionais. Para tratar esta integral, Pankratz et al. (2005) usaram a aproximação de Laplace, como definido em Ripatti & Palmgren (2000), o que viabilizou a obtenção das estimativas dos parâmetros β e dos componentes do vetor g . Este processo de estimação foi implementado na função *coxme* no pacote *kinship* (THERNEAU, 2007; R DEVELOPMENT CORE TEAM, 2011).

Para testar se o efeito das covariáveis, representados pelos parâmetros β , são significativos, foi utilizado o teste de Wald (WALD, 1943), o qual considera as hipóteses: $H_0: \beta_p = 0$ versus $H_a: \beta_p \neq 0$. Se a hipótese nula for rejeitada, haverá evidências de efeito significativo, indicando que a covariável X_p deve ser mantida no modelo. No que diz respeito aos componentes do vetor g , deve-se testar a significância dos componentes de variância σ_g^2 e σ_f^2 . Assim, se $\Sigma = 2\phi\sigma_g^2$ deve-se testar as hipóteses $H_0: \sigma_g^2 = 0$ versus $H_a: \sigma_g^2 \neq 0$. Rejeição da hipótese nula implicará em efeito genético (poligênico) significativo. Dada às dificuldades em se estabelecer uma estatística de teste para testar as hipóteses associadas aos componentes de variância σ_g^2 e σ_f^2 , intervalos de confiança foram calculados para essa finalidade. Tais intervalos foram obtidos com base em um método que faz uso da função de verossimilhança perfilada. Esse método basicamente reduz o logaritmo da função de verossimilhança para uma função do parâmetro de interesse tratando todos os demais como parâmetros *nuisance* (de perturbação) e, então, maximizando sobre eles. Para mais detalhes sobre este método ver Venzon e Moolgavkar (1998). Programa em R utilizado para obtenção desses intervalos encontra-se no ANEXO - 1 deste trabalho.

Quanto à verificação do pressuposto de riscos proporcionais e da qualidade de ajuste global do modelo de Cox de efeitos mistos, pesquisas

adicionais são necessárias, pois atualmente não existem metodologias específicas na literatura propostas para essa finalidade. Para tal, foi assumido que, se o pressuposto de riscos proporcionais for atendido para as covariáveis no modelo de Cox usual, este também será válido para o modelo de Cox de efeitos mistos.

3. RESULTADOS

3.1 ANÁLISE DESCRITIVA

Com o objetivo de identificar se fatores genéticos e ou ambientais estão associados com a idade em que a hipertensão arterial se manifesta foi realizada, inicialmente, uma análise descritiva dos dados descritos na Seção 2.1. Para esses dados, a variável resposta é composta pelo par (t_i, δ_i) , sendo t_i a idade de diagnóstico da hipertensão e δ_i a variável indicadora tal que, indivíduos hipertensos assumem valor 1 e normotensos valor 0.

Foram considerados na análise 1673 indivíduos, sendo 35,6% classificados como hipertensos (falhas) e 64,4% como normotensos (censuras). Estimativa da curva de sobrevivência desses indivíduos, obtida pelo estimador de Kaplan-Meier sem levar em conta as covariáveis e a estrutura familiar, é apresentada na Figura 2. Desta figura, pode-se observar que $S(t) = 0,50$ para $t \approx 57$, indicando que o tempo mediano de diagnóstico de hipertensão foi de 57 anos.

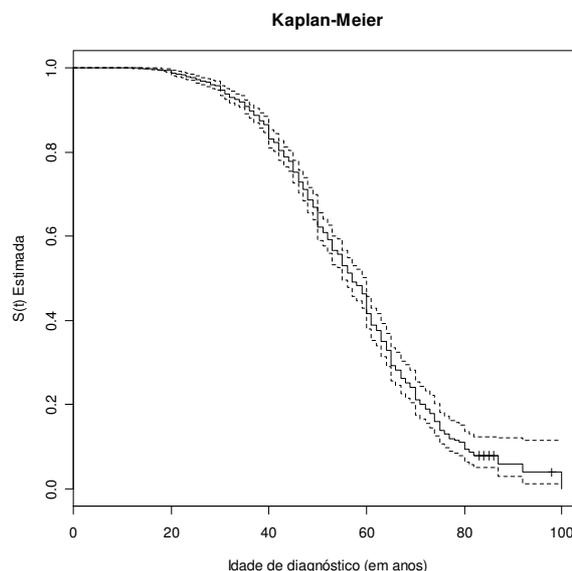


FIGURA 2: Sobrevivência estimada por Kaplan-Meier. Os tempos representados por + mostram onde ocorreram censuras.

As covariáveis eleitas pelos pesquisadores do InCor, apresentadas na Tabela 1, foram dicotomizadas com o auxílio e interesse dos mesmos para a incorporação destas no ajuste dos modelos. As frequências e a prevalência de

hipertensão, de acordo com as categorias destas covariáveis, estão mostradas na Tabela 2. A partir desta tabela, pode ser destacado um maior percentual de indivíduos do gênero feminino, 56%; prevalência de hipertensão de 52% para indivíduos com IMC acima do normal (valores maiores que 25) e de 63% para indivíduos com circunferência abdominal acima do normal (igual ou menor que 94cm para mulheres e 100cm para homens), bem como 68% de indivíduos classificados como sedentários, isto é, que declararam não praticar atividade física nenhum dia da semana.

TABELA 2: Frequências e prevalência de hipertensão de acordo com as covariáveis dicotomizadas.

Variáveis	Categoria	Frequência	Percentual	Prevalência
Gênero	Masculino	728	44%	33%
	Feminino	945	56%	37%
IMC	Até 25	1.005	60%	25%
	Acima de 25	632	38%	52%
	NA	36	2%	-
Sedentarismo	Sedentário	1.142	68%	37%
	Não sedentário	528	32%	32%
	NA	3	0%	-
Circunferência Abdominal	Normal	1.301	78%	28%
	Acima do Normal	353	21%	63%
	NA	19	1%	-

IMC: Índice de massa corporal; NA: Dados faltantes.

As curvas de sobrevivência apresentadas na Figura 3 e os testes *logrank* mostrados na Tabela 3, realizados para compará-las, indicam que tais curvas para as categorias das covariáveis gênero, IMC e circunferência abdominal diferem entre si (valor $p < 0,0001$).

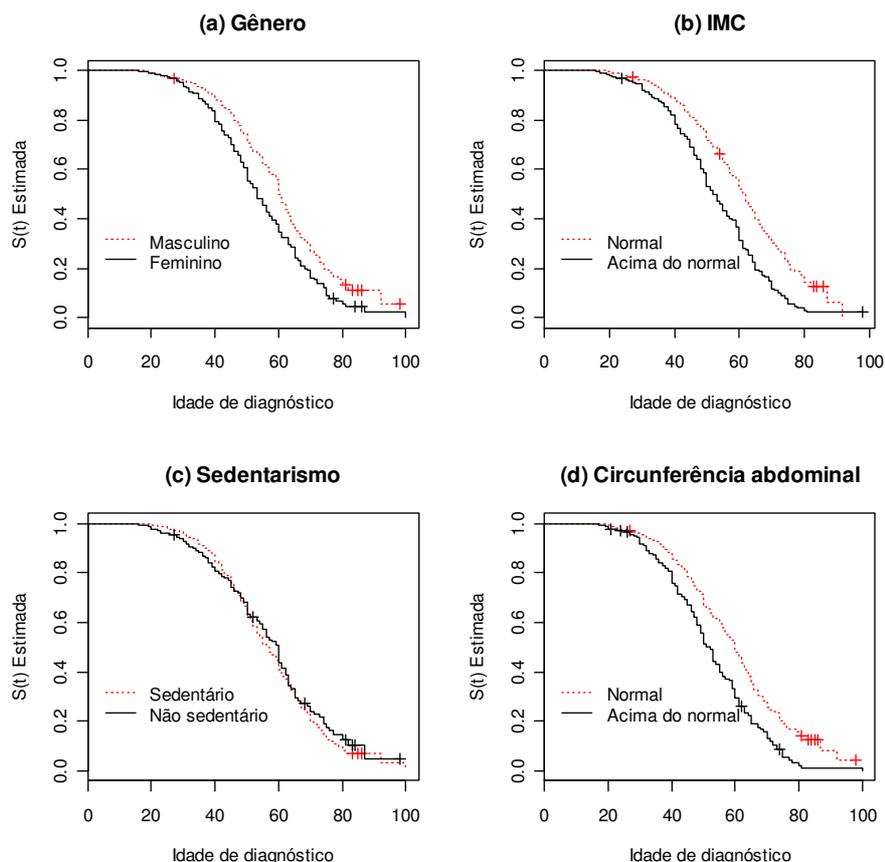


FIGURA 3: Curvas de sobrevivência estimadas por Kaplan-Meier. Os tempos representados por + mostram onde ocorreram censuras em cada grupo.

Observando as curvas na Figura 3a, tem-se que os indivíduos do sexo masculino foram diagnosticados com hipertensão em idades mais avançadas do que os do sexo feminino. Este fato pode provavelmente ser decorrente do fato de que, em geral, os homens vão menos regularmente aos médicos do que as mulheres. Também foram diagnosticados com hipertensão em idades mais avançadas, indivíduos com IMC até 25, bem como os com circunferência abdominal normal (Figuras 3b e 3d). Já quanto aos indivíduos sedentários e não-sedentários neste estudo, estes apresentaram diagnóstico de hipertensão em idades similares (Figura 3c).

TABELA 3: Resultados do teste logrank utilizado para comparar as curvas de sobrevivência obtidas para as categorias de cada covariável considerada.

Covariável	Estatística de teste	Valor p
Gênero	25,6	< 0,0001 *
IMC	48,1	< 0,0001 *
Sedentarismo	0,01	0,78
Circunferência abdominal	36,8	< 0,0001 *

*Valores significativos a 5%.

Com a finalidade de evitar problemas relacionados à multicolinearidade no ajuste de modelos, foi utilizado o teste qui-quadrado de Pearson para identificar covariáveis correlacionadas. Como resultado, foi constatado alta correlação entre as covariáveis IMC e circunferência abdominal (valor p menor do que 0,0001). Sendo tal correlação esperada, uma vez que indivíduos com IMC alto tendem a ter circunferência abdominal maior, optou-se pela utilização da covariável IMC nos modelos, por ser a mais comumente utilizada pelos profissionais da saúde.

3.2 AJUSTE DE MODELOS

Utilizando primeiramente o modelo de Cox para análise dos dados, foram ajustados no R os modelos sequenciais possíveis considerando as covariáveis eleitas na Seção 3.1, isto é IMC e Gênero. Os resultados estão na Tabela 4.

TABELA 4: Estimativas obtidas para os parâmetros nos modelos de Cox ajustados

Modelos	Estimativas	Valor P
1 - IMC	$\hat{\beta}_1=0,568$	< 0,0001 *
2 – IMC + Gênero	$\hat{\beta}_1=0,508$	< 0,0001 *
	$\hat{\beta}_2= 0,336$	< 0,0001 *
3 – IMC+ Gênero + Gênero*IMC	$\hat{\beta}_1= 0,432$	0,0011 *
	$\hat{\beta}_2= 0,266$	0,0370 *
	$\hat{\beta}_3=0,132$	0,45

*Valores significativos a 5%.

A partir dos resultados apresentados na Tabela 4, tem-se que a interação entre IMC e gênero não é significativa ($p = 0,45$), mas que os efeitos principais de IMC e gênero são significativos ($p < 0,05$). Em consequência, a avaliação do pressuposto de riscos proporcionais foi realizada utilizando-se os resíduos padronizados de Schoenfeld associados ao modelo 2. Tais resíduos, para cada covariável no modelo, podem ser visualizados na Figura 4.

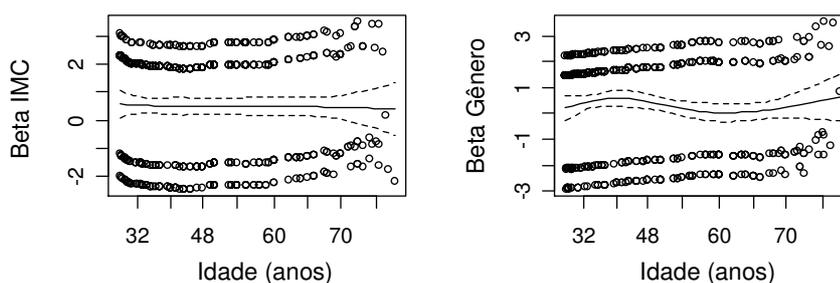


FIGURA 4: Resíduos padronizados de Schoenfeld do modelo de Cox com as covariáveis IMC e gênero

Com base na Figura 4, é possível observar inclinações próximas de zero, ou seja, a ausência de tendências acentuadas, o que pode ser confirmado pelos valores de ρ próximos de zero e respectivos valores $p \gg 0,05$ associados ao teste de proporcionalidade dos riscos mostrados na Tabela 5. Sendo assim, não há evidências para rejeição da hipótese nula de proporcionalidade dos riscos.

TABELA 5: Testes de proporcionalidade dos riscos no modelo de Cox

Covariável	$\rho(\rho)$	χ^2	Valor P
IMC	-0,0193	0,214	0,644
Gênero	-0,0287	0,468	0,494
GLOBAL	-	0,802	0,664

Embora as covariáveis IMC e gênero tenham apresentado efeitos significativos no modelo de Cox e o pressuposto de riscos proporcionais não tenha sido violado, extensão do modelo de Cox, denominado modelo de Cox de efeitos mistos (descrito na Seção 2.2.2) foi utilizado a fim de considerar a estrutura de correlação existente nos dados e, assim, estimar os componentes

de variância relacionados aos fatores genéticos e ambientais. Para o ajuste desse modelo foi utilizada a função *coxme* do pacote *kinship* do R. Os resultados estão na Tabela 6.

TABELA 6: Estimativas obtidas para os parâmetros nos modelos de Cox de Efeitos Mistos

IMC + Gênero	Estimativas	Valor p	$\hat{\sigma}_g^2$	$\hat{\sigma}_f^2$	Log-verossimilhança
Modelo 1 - Sem efeito aleatório	$\hat{\beta}_1=0,544$	<0,0001*			-3445,063
	$\hat{\beta}_2= 0,360$	<0,0001*			
Modelo 2 – Com efeito poligênico	$\hat{\beta}_1=0,586$	<0,0001*	0,744		-3302,381
	$\hat{\beta}_2= 0,421$	<0,0001*			
Modelo 3 – Com efeito poligênico e ambiental	$\hat{\beta}_1= 0,577$	<0,0001*	0,679	0,039	-3308,24
	$\hat{\beta}_2= 0,423$	<0,0001*			

*Valores significativos a 5%.

O teste de Wald foi utilizado para verificar a significância dos parâmetros β , sendo que os três modelos apresentados na Tabela 6 apresentam valores p significativos a 5%. Já para testar a significância dos parâmetros $\hat{\sigma}_g^2$ no modelo 2 e $\hat{\sigma}_g^2$ e $\hat{\sigma}_f^2$ no modelo 3, intervalos de confiança foram obtidos utilizando-se o conceito apresentado na Seção 2.2.3. Tais intervalos podem ser visualizados nas Figuras 5 e 6.

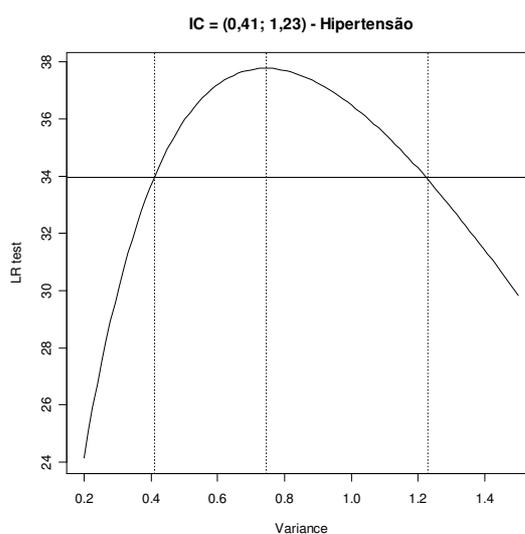


FIGURA 5: Representação gráfica do intervalo de confiança associado ao parâmetro σ_g^2 efeito poligênico do modelo 2.

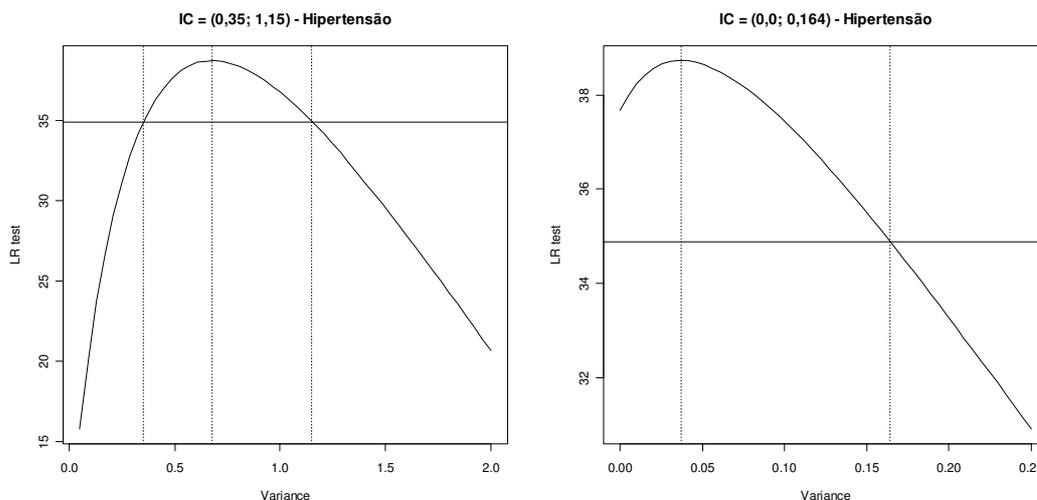


FIGURA 6: Representação gráfica dos intervalos de confiança associados aos parâmetros σ_g^2 e σ_f^2 , efeitos poligênico e ambiental do modelo 3.

Os intervalos de confiança para o modelo 3 são $IC(\sigma_g^2, 95\%) = (0,35; 1,15)$ e $IC(\sigma_f^2, 95\%) = (0,0; 0,164)$, neste último, o valor 0 está contido no intervalo, o que explicita que o parâmetro de efeito ambiental σ_f^2 é não significativo. Já o ajuste do modelo 2 apresenta o efeito poligênico σ_g^2 significativo, $IC(\sigma_g^2, 95\%) = (0,41; 1,23)$, desta forma, decidiu-se pela escolha deste modelo e as seguintes interpretações são obtidas a partir da Tabela 7.

TABELA 7: Resultados do ajuste do modelo de Cox de Efeitos Mistos

Efeito	Estimativa	$IC_{95\%}$ (Estimativa)	RR	$IC_{95\%}$ (RR)
IMC	0,586	(0,376;0,795)	1,796	(1,456;2,214)
Gênero	0,421	(0,209;0,632)	1,523	(1,232;1,881)
σ_g^2	0,744	(0,41;1,23)	2,363	(1,506;3,421)

IC: intervalo de 95% de confiança; RR: razão de riscos.

O risco de indivíduos com IMC acima de 25 desenvolver hipertensão em uma idade t (mantendo-se fixa a covariável gênero e o efeito aleatório) é 79% maior do que o risco de indivíduos com IMC abaixo deste valor. Com um intervalo de confiança de 95%, pode-se dizer que este risco varia entre 45 e 121%. Desse modo, para os indivíduos desse estudo, pode-se concluir que o

risco da hipertensão se manifestar em idades inferiores é maior dentre aqueles indivíduos com $IMC \geq 25$ do que entre os com $IMC < 25$. Já quanto ao risco de indivíduos do gênero feminino desenvolver hipertensão em uma específica idade t (mantendo fixa a covariável IMC e o efeito aleatório), este foi 52% maior do que o risco dos indivíduos do gênero masculino. Com um intervalo de confiança de 95%, pode-se dizer que este risco varia entre 23 e 88%. Sendo assim, para os indivíduos nesse estudo, o risco da hipertensão se manifestar em uma específica idade t foi indicado ser maior entre as mulheres do que entre os homens. Contudo, e como já mencionado anteriormente, isto pode ser decorrente do fato dos homens, em geral, procurarem por médicos menos frequentemente do que as mulheres havendo, assim, um diagnóstico mais tardio da hipertensão entre os homens.

Como no modelo selecionado o efeito poligênico foi significativo, uma vez que o componente de variância poligênico apresentou estimativa de 0,744 com intervalo de 95% de confiança de 0,41 e 1,23, pode-se concluir que efeitos genéticos também influenciam na idade em que a hipertensão se manifesta. Note que significância do componente genético implica na existência de variabilidade genética significativa entre as famílias, sugerindo que fatores genéticos também estão associados à idade em que a hipertensão se manifesta. Logo, como $\exp(\sqrt{0,744}) = 2,36$, pode-se concluir que existem indivíduos nesse estudo que apresentam risco de desenvolver hipertensão em uma idade t o qual é 136% maior do que o risco médio dos indivíduos, sendo tal risco devido a efeitos genéticos (ou seja, transmitido de geração em geração).

A partir do ajuste do modelo 2, foram obtidas as estimativas dos efeitos genéticos g_{ij} associados aos indivíduos i ($i = 1, \dots, n_j$) de cada família j ($j = 1, \dots, 81$). Tais estimativas estão mostradas (por família) nas Figuras 7 e 8, sendo possível visualizar as variabilidades das mesmas entre os indivíduos e entre as

famílias.

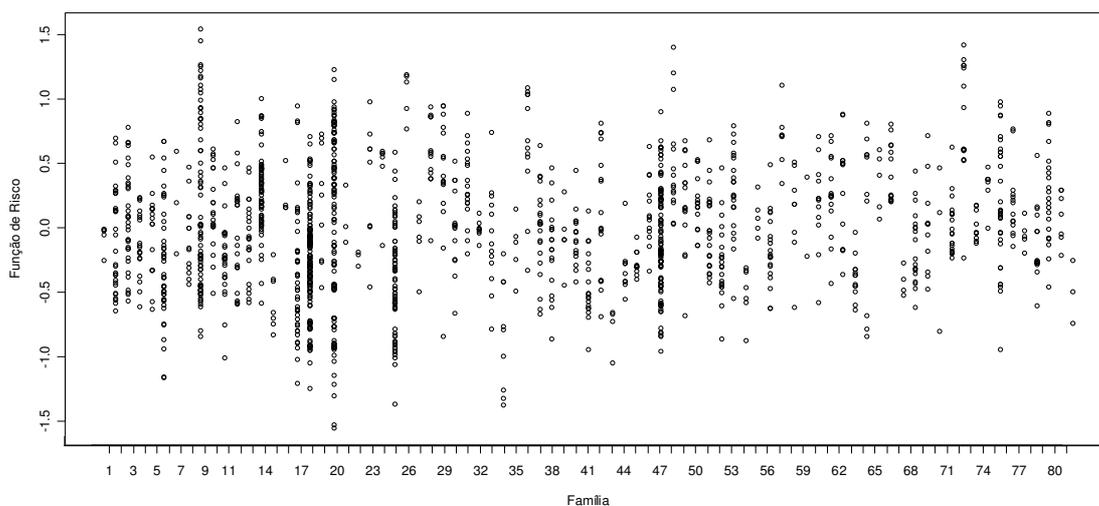


FIGURA 7: Estimativas dos efeitos genéticos g_{ij} associados aos indivíduos i ($i = 1, \dots, n_j$) de cada família j ($j = 1, \dots, 81$).

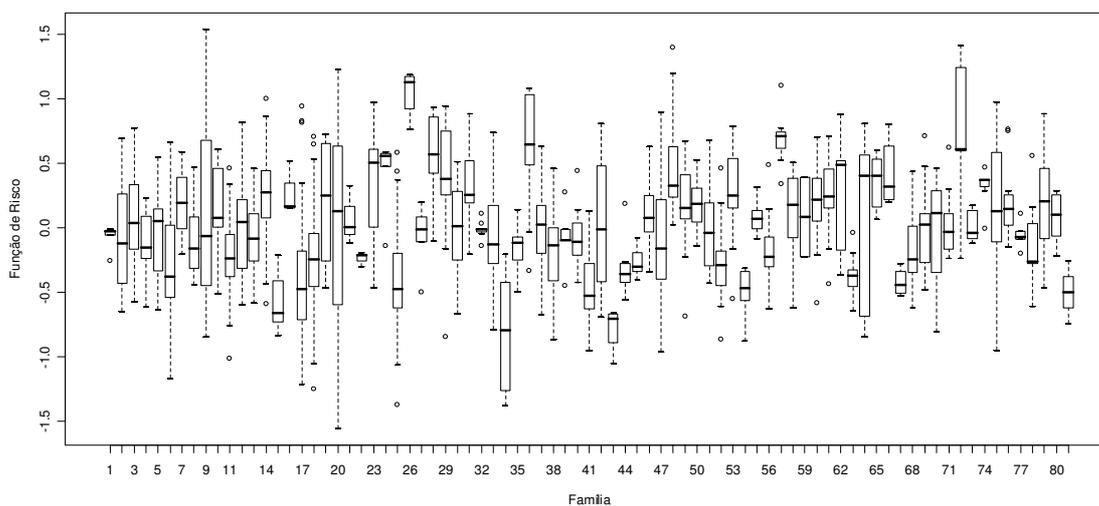


FIGURA 8: Boxplots das estimativas dos efeitos genéticos g_{ij} associados aos indivíduos i ($i = 1, \dots, n_j$) de cada família j ($j = 1, \dots, 81$).

Com base nas estimativas dos efeitos genéticos g_{ij} e da interpretação do componente de variância σ_g^2 , duas famílias foram escolhidas a fim de ilustrar e exemplificar as conclusões obtidas nesse estudo. Nesse sentido, as Figuras 9 e 10 apresentam, respectivamente, os indivíduos das famílias 48 e 45, com as respectivas idades de diagnóstico de hipertensão para os hipertensos (círculos ou quadrados preenchidos) e idades dos indivíduos normotensos no momento do estudo, bem como o risco de hipertensão (isto é, $\exp(g_{ij})$) devido a efeitos genéticos.

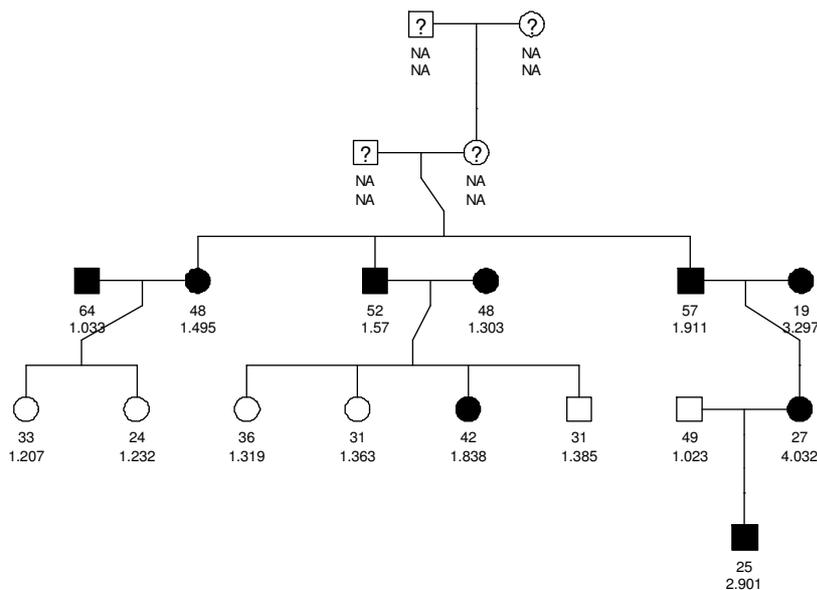


FIGURA 9: Árvore genealógica dos indivíduos da família 48, com suas respectivas idades de diagnóstico de hipertensão (ou na data de pesquisa para normotensos) e risco de hipertensão

A partir da Figura 9 é possível observar que o risco de hipertensão devido a fatores genéticos é maior na família 48, pois os indivíduos desta família apresentam riscos maiores de desenvolver hipertensão. Também é possível observar que a idade de sua manifestação é uma tendência transmitida pelos familiares. Conseqüentemente, o risco de hipertensão associado ao efeito poligênico da família 45 (Figura 10) é menor, evidenciando que os indivíduos desta família têm risco menor de desenvolver hipertensão.

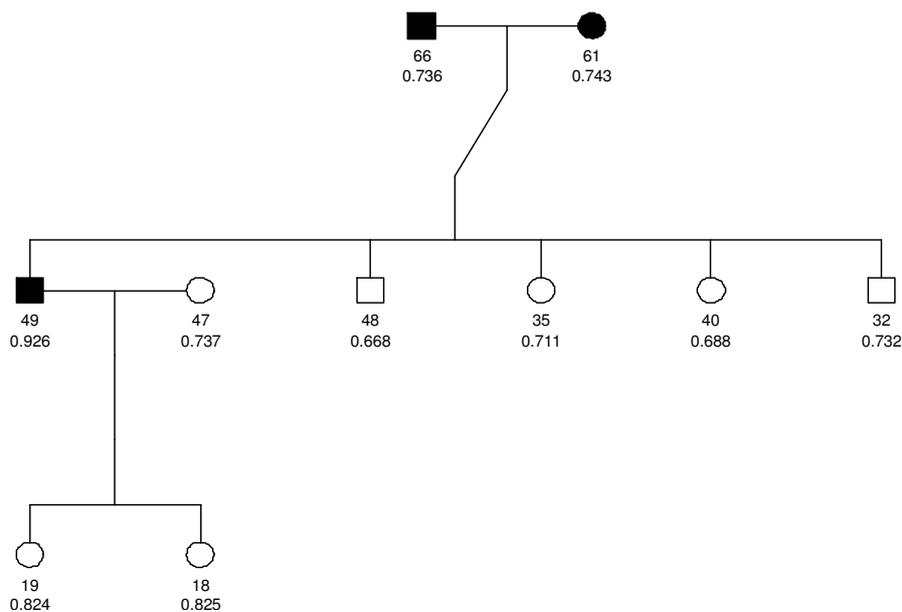


FIGURA 10: Árvore genealógica dos indivíduos da família 45, com suas respectivas idades e risco de hipertensão.

Para visualizar graficamente os riscos dos indivíduos das famílias 48 e 45, são apresentadas na Figura 11 as estimativas das funções de risco acumulada associadas aos indivíduos das famílias 48 (em vermelho) e 45 (em azul), juntamente com a função de risco acumulada de um indivíduo com efeito poligênico nulo, $g_{ij} = 0$ (em preto). A partir desta figura, fica evidenciado que os indivíduos da família 45 apresentam risco menor de hipertensão ao longo do eixo da idade quando comparados aos indivíduos da família 48. Ainda, é possível observar uma variabilidade maior entre os indivíduos da família 48 do que entre os da família 45, o que justifica a significância estatística do componente de variância poligênico.

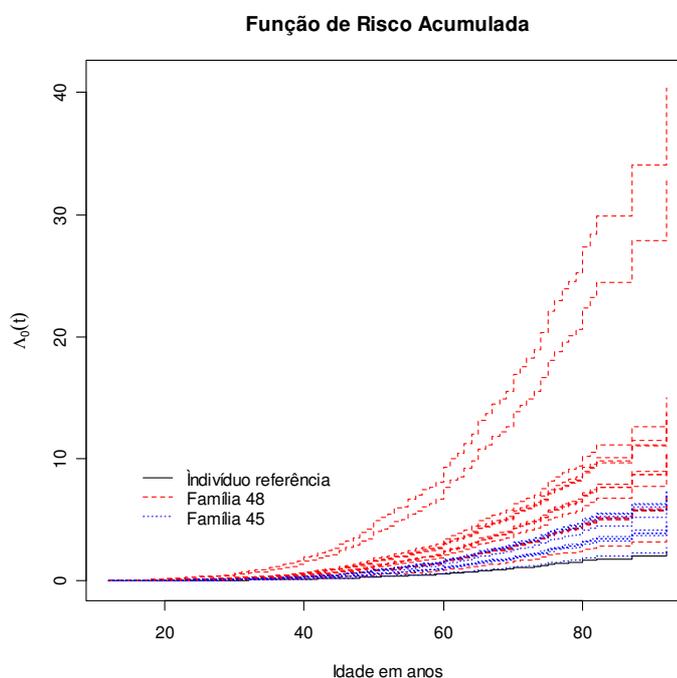


FIGURA 11: Representação das funções de risco acumulada estimadas para os indivíduos das famílias 48 (em vermelho) e 45 (em azul) e para um indivíduo com efeito genético $g_{ij} = 0$.

4. CONCLUSÃO E DISCUSSÃO

Neste trabalho, dados de um estudo sobre hipertensão e fatores de risco (ou covariáveis) associados a ela foi analisado. Com intuito exploratório foi, inicialmente, realizada uma análise descritiva dos dados, bem como ajustado o modelo de Cox com as covariáveis IMC e gênero. Tal modelo, embora tenha apresentado ajuste satisfatório e a suposição de riscos proporcionais não tenha sido violada, não se apresentava como o mais adequado, pois os dados foram coletados em famílias de modo que a estrutura de correlação dos indivíduos exigia um tratamento diferenciado. Desse modo, para que a informação do parentesco existente entre os indivíduos de uma mesma família pudesse ser considerada, bem como a questão principal do estudo sobre a associação de fatores genéticos e ambientais com a idade em que a hipertensão arterial se manifesta pudesse ser respondida, foi utilizado o modelo de Cox de efeitos mistos. Com base nesse modelo, os componentes de variância associados aos efeitos poligênico e ambiental foram estimados, sendo que somente o efeito poligênico se mostrou significativo. Quanto à adequação do modelo de Cox de efeitos mistos ajustado, foi assumido que o pressuposto de riscos proporcionais, avaliado no modelo de Cox usual, se estendia a este modelo. Contudo, estudos adicionais nesse sentido são importantes e necessários.

De modo geral, os resultados obtidos apontaram que as covariáveis IMC e gênero, bem como fatores genéticos, estão associados à hipertensão e à idade em que esta se manifesta. Tais conclusões levam à algumas considerações para o direcionamento e andamento desse estudo, pois se fatores genéticos estão associados à idade em que a hipertensão se manifesta, a questão que se segue passa a ser a de identificar os genes responsáveis. Nesse sentido, abre-se uma nova fase de pesquisa do Estudo Corações de Baependi, MG, em que o genoma destes indivíduos será mapeado a fim de se buscar identificar quais genes são responsáveis pelo desenvolvimento da doença.

Nesse contexto, amostras do genoma humano têm sido utilizadas, pois o sequenciamento do genoma ainda é muito caro em termos financeiros. Tais

amostras, contudo, compreendem uma quantidade grande de informações de cada indivíduo (da ordem de um milhão ou mais de locos ou marcadores do genoma). Desse modo, abre-se aí um leque de oportunidades para o desenvolvimento de novas metodologias estatísticas, que com o apoio de áreas como a bioinformática e a genética poderão, no futuro, trazer avanços e descobertas importantes para auxiliar nas políticas públicas e de saúde da população.

REFERÊNCIAS

- AALEN, O.O. (1978). Nonparametric Inference for a Family of Counting Processes. **Annals of Statistics**, (6), 701-726.
- AVEZUN, A.; GUIMARÃES, H.P. (2010). Hipertensão Arterial Sistólica. **Brazilian Journal of Hypertension**, v. 17, (1,2): 7-10.
- BRESLOW, N. (1972). Contribuição à Distribuição do Artigo COX, D.R.. **Journal of the Royal Statistical Society B**, **34**, 216-217.
- COLOSIMO, E.A; GIOLO, S.R. (2006). **Análise de Sobrevida Aplicada**. São Paulo: Edgard Blucher, 390p.
- COX, D.R. (1972). Partial Linkelihood. *Biometrika*, **62**, 269-76.
- DUCROCQ, V.; CASSELA, G. (1996) A Bayesian analysis of mixed survival models, **Genet. Sel. Evol.** 28, pp. 505–529.
- GIOLO, S.R; PEREIRA, A.C.; DE ANDRADE, M.; OLIVEIRA, C.M.; KRIEGER, J.E.; SOLER, J.M.P. (2009). Genetic Analysis of Age-at-Onset for Cardiovascular Risk Factors in a Brazilian Family Study. **Human Heredity**, 68: 131-138.
- KAPLAN, E. L.; MEIER, P. (1958). Nonparametric estimation from incomplete observations. **Journal of the American Statistical Association** 53, 457-48 1.
- KULLO, I.J.; TURNER, S.T.; KARDIA, S.L.R.; MOSLEY, T.H.; BOERWINKLE, E.; DE ANDRADE, M. (2006). A Genome-Wide Linkage Scan for Ankle-Branchial Index in African American and non-Hispanic White Subjects Participating in the GENOA Study. **Atherosclerosis**, 187 (2): 433-438.
- KULLO, I.J.; DING, K. (2007). Mechanisms of Disease: the genetic basis of coronary heart disease. **Nature Clinical Practice Cardiovascular Medicine**, **4**: 558-569.
- MAHANEY, M.C.; BLANGERO, J.; COMUZZIE, A.G.; VANDEBERG, J.L.; STERN, M.P.; MCCLUER, J.W. (1995). Plasma HDL cholesterol, triglycerides, and adiposity. A Quantitative Genetic Test of the Conjoint Trait Hypothesis in the San Antonio Family Heart Study. **Circulation**, 92: 3240-3248.
- MANTEL, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. **Cancer Chemotherapy Reports**, 50, 163-170.
- OLIVEIRA, C.M.; PEREIRA, A.C.; DE ANDRADE, M.; SOLER, J.M.P.; KRIEGER, J.E. (2008). Heritability of cardiovascular risk factors in Brazilian population: Baependi Heart Study. **BMC Medical Genetics**, 9:32 doi:10.1186/1471-2350-9-32.
- PANKRATZ, V.S; DE ANDRADE, M.; THERNEAU, T. (2005). Random-effects Cox proportional hazards model: general variance components methods for time-to-event data. **Genetic Epidemiology**; 28: 97-109.

PENA, S.D. (2000). O Retrato Molecular do Brasil. **Ciência Hoje**, 27 (159): 16-25.

PIMENTA, J.R.; ZUCCHERATO, L.W.; DEBES, A.A.; MASELLI, L.M.F.; SOARES, R.P.S.; MOURA-NETO, R.S.; ROCHA, J.; BYDLOWSKI, S.P.; PENA, S.D. (2006). Color and Genomic Ancestry in Brazilians: A Study with Forensic Microsatellites. **Human Heredity**, 62: 190-195.

PEREIRA, A.C.; SPOSITO, G.F.; MOTA, G.F.; CUNHA, R.S.; HERKENHOFF, F.L.; (2006). Endothelial nitric oxide synthase gene variant modulates the relationship between serum cholesterol levels and blood pressure in the general population: new evidence for a direct effect of lipids in arterial blood pressure. **Atherosclerosis**; 184:193-200.

R DEVELOPMENT CORE TEAM (2011). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

RIPATTI, S.; PALMGREN J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. **Biometrics**; 56:1016-1022.

THERNEAU, T. (2007). **On mixed-effect Cox models, sparse matrices, and modeling data from large pedigrees**. New York. 54p. Relatório técnico.

THOMAS,D.C.; HAILE, R.W. et al. (2005). Recent Developments in Genomewide Association Scans: A workshop summary and review. **Am. J. Hum. Genet.**, 77: 337-345.

VEZON, D.J.; MOOLGAVKAR, S.H. (1988). A Method for Computing Profile-Likelihood-Based Confidence Intervals. **Journal of the Royal Statistical Society**. Series C (Applied Statistics), 37: 87-94.

WALD, A. (1943). Tests of Statistical Hypotheses concerning Several Parameters when the Number of Observations is Large. **Trans Amer Math. Soc.**, 54, 426, 482.

ANEXO

ANEXO 1 – Obtenção da Figura 5 – Intervalo de confiança

```

fit0<-coxph(Surv(idpa,censpa)~ factor(BMI_25)+ factor(SEX))
  fit1<-coxme(Surv(idpa,censpa)~factor(BMI_25)+  factor(SEX),  random=
~1|ID, varlist=kmat)
  vtemp <- seq(0.01, 1.5, length=100)
  ltemp <- 0*vtemp
  for (i in 1:length(vtemp)) {
    tfit<-coxme(Surv(idpa,censpa)~factor(BMI_25)+  factor(SEX),  random=
~1|ID, varlist=kmat, variance= vtemp[i])
    ltemp[i] <- -2*(fit0$loglik[2]-tfit$loglik[2])
  }
  plot(vtemp, ltemp, xlab='Variância', ylab="LR test", type="l")
  abline(h=-2*(fit0$loglik[2]-fit1$loglik[2]) - 3.84)
  abline(v=0.744462,lty=3)
  abline(v=0.41,lty=3)
  abline(v=1.23,lty=3)
  title("IC = (0,41; 1,23)")

```