

**MODELAGEM E ANÁLISE DE FATORES  
INFLUENTES NA GRAVIDEZ NA  
ADOLESCÊNCIA  
NA MICRORREGIÃO DE CAMPO BELO-MG**

**ALADIR HORACIO DOS SANTOS**

**2009**

**ALADIR HORACIO DOS SANTOS**

**MODELAGEM E ANÁLISE DE FATORES INFLUENTES NA  
GRAVIDEZ NA ADOLESCÊNCIA  
NA MICRORREGIÃO DE CAMPO BELO-MG**

Tese apresentada à Universidade Federal de Lavras,  
como parte das exigências do Programa de Pós-  
graduação em Estatística e Experimentação  
Agropecuária, para obtenção do título de "Doutor".

Orientador

Prof.Dr. Augusto Ramalho de Moraes

LAVRAS  
MINAS GERAIS - BRASIL  
2009

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da  
Biblioteca Central da UFLA**

Santos, Aladir Horácio.

Modelagem e análise de fatores influentes na gravidez na  
adolescência na microrregião de Campo Belo, MG / Aladir Horácio  
Santos. – Lavras : UFLA, 2009.

124 p. : il.

Tese (doutorado) – Universidade Federal de Lavras, 2009.

Orientador: Augusto Ramalho de Moraes.

Bibliografia.

1. Gravidez na adolescência. 2. Saúde pública. 3. Análise de  
sobrevivência. 4. Modelos de Cox. I. Universidade Federal de  
Lavras. II. Título.

CDD – 519.536

**ALADIR HORACIO DOS SANTOS**

**MODELAGEM E ANÁLISE DE FATORES INFLUENTES NA  
GRAVIDEZ NA ADOLESCÊNCIA  
NA MICRORREGIÃO DE CAMPO BELO-MG**

Tese apresentada à Universidade Federal de Lavras,  
como parte das exigências do Programa de Pós-  
graduação em Estatística e Experimentação  
Agropecuária, para obtenção do título de "Doutor".

APROVADA em 10 de julho de 2009

Prof. Dr. Denismar Alves Nogueira	UFAL
Prof. Dr. Marcelo Silva de Oliveira	UFLA
Prof. Dr. Mario Javier Ferrua Vivanco	UFLA
Prof. Dr. Washington Santos da Silva	CEFET-MG
Prof. Dr. José Barbosa Júnior	UNIFOR-MG

Prof. Dr. Augusto Ramalho de Moraes  
UFLA  
(Orientador)

LAVRAS  
MINAS GERAIS - BRASIL

*À memória de Maria Ana Ramos, “Mãe Dinha”*

**Dedico**

*À minha esposa, Neide A. Ferreira Santos*

*Aos meus filhos, Alexandre Augusto e Ana Cristina Ferreira Santos*

**Ofereço**

## AGRADECIMENTOS

À Universidade Federal de Lavras e ao Departamento de Ciências Exatas, por ter-me dado a oportunidade de crescer.

Ao professor Augusto Ramalho de Moraes, pela orientação.

Ao professor Mario Javier Ferrua Vivanco, pela coorientação, paciência e amizade.

Aos professores do Departamento de Ciências Exatas da UFLA, pelos ensinamentos e experiências transmitidos.

Ao Diretor-Administrativo do Hospital da Santa Casa de Misericórdia de Campo Belo, Anataniel Reis Oliveira, pela disponibilização dos dados, sem os quais esse trabalho não seria possível.

Ao Governo do Estado de Minas Gerais pela concessão da licença remunerada.

As enfermeiras e auxiliares de enfermagem dos PSF's.

À professora Rosália Vilela, diretora da Escola Estadual Padre Alberto Fuger, pela compreensão e incentivo.

À Regina e a Cidinha, Bibliotecárias do UNIFOR-MG, pela ajuda valiosa.

Aos colegas de curso pela amizade.

A DEUS, por proporcionar-me essa oportunidade de crescimento e por ter protegido minha vida durante esse período.

Enfim, a todos que contribuíram para realização deste trabalho.

## SUMÁRIO

	<b>Página</b>
LISTA DE FIGURAS.....	i
LISTA DE TABELAS.....	iv
RESUMO.....	viii
ABSTRACT.....	ix
1 INTRODUÇÃO.....	1
2 REFERENCIAL TEÓRICO.....	3
2.1 Gravidez na adolescência.....	3
2.2 Análise de sobrevivência.....	13
2.2.1 Censura.....	13
2.3 Inferência não-paramétrica.....	14
2.4 Modelos de regressão.....	15
2.4.1 Variáveis explanatórias.....	15
2.4.2 Formulação de um modelo de regressão.....	16
2.5 Modelos de regressão.....	18
2.6 Modelos de riscos proporcionais de Cox.....	18
2.7 Estimação dos parâmetros no modelo de Cox.....	20
2.8 Estimação de funções relacionadas com $\alpha_o(t)$ .....	22
2.8.1 Interpretação dos parâmetros no modelo de Cox.....	23
2.9 Suposição de riscos proporcionais no modelo de Cox.....	24
2.10 Generalização do modelo de regressão de Cox.....	24
2.11 Modelo de Cox estratificado.....	25
2.12 O modelo sem interação.....	27
2.13 Modelo de Cox estendido.....	34
2.14 Adequação do ajuste do modelo de Cox.....	41
2.15 Modelos de regressão paramétricos.....	43

2.16 Modelo de regressão exponencial.....	45
2.17 Modelo de regressão Weibull.....	46
2.18 Modelo de regressão Log-normal.....	46
2.19 Modelo de regressão Gama generalizada.....	47
2.20 Inferência nos modelos de regressão paramétrico.....	48
3 MATERIAL E MÉTODOS.....	50
3.1 Material.....	50
3.2 Método.....	52
3.2.1 Definição do evento.....	52
3.2.2 Descrição das covariáveis.....	53
3.3 Análise dos dados.....	54
4 RESULTADOS E DISCUSSÃO.....	57
4.1 Análise descritiva e exploratória dos dados.....	57
4.2 Seleção do modelo paramétrico.....	58
4.3 Adequação do modelo.....	71
4.4 Interpretação das estimativas.....	76
4.5 Análise no modelo de regressão de Cox.....	79
4.6 Interpretação das estimativas.....	85
4.7 Seleção, avaliação e adequação do modelo de Cox - riscos propor- cionais.....	88
4.8 Modelo de Cox estratificado.....	96
4.9 Interpretação das estimativas das covariáveis no modelo Cox estratificado.....	104
4.10 Modelo de Cox com covariáveis dependentes do tempo.....	105
5 CONCLUSÕES.....	116
6 REFERÊNCIAS BIBLIOGRÁFICAS.....	117
ANEXO.....	123



## LISTA DE FIGURAS

	<b>Página</b>
FIGURA 1	Representação gráfica das curvas de sobrevivência estimadas pelo método de Kaplan & Méier (1958) para a covariável religião católica e evangélicas .....57
FIGURA 2	Representações gráficas das análises dos resíduos de Cox-Snell para os modelos log-normal, exponencial e weibull ajustados aos dados. (Continua...)..... 73
FIGURA 3	Representações gráficas da sobrevivência dos resíduos estimados por K-M e pelo modelo Log-Normal e da curva sobrevivência estimada. .... 75
FIGURA 4	Representações gráficas das curvas de sobrevivência dos resíduos estimados por K-M e pelo modelo Exponencial padrão e da curva sobrevivência estimada. .... 76
FIGURA 5	Representações gráficas do logaritmo do risco $Log(\hat{\Lambda}_{0j}(t))$ em relação ao tempo em dias após a gravidez para as covariáveis religião (Rel), renda (Ren), programas sociais (OS) e vida social ativa (VSA)..... 90
FIGURA 6	Representações gráficas do logaritmo do risco $Log(\hat{\Lambda}_{0j}(t))$ em relação ao tempo em dias após a gravidez para as covariáveis orientação sexual (OS), responsabilidade legal do parceiro (RL), vida social ativa (VSOA) e escolaridade da gestante (E)..... 91
FIGURA 7	Representação gráfica da suposição de riscos proporcionais para a covariável renda (Ren)..... 92
FIGURA 8	Representação gráfica da suposição de riscos proporcionais para a religião (Rel)..... 92

FIGURA 9	Representação gráfica da suposição de riscos proporcionais para a covariável responsabilidade legal do parceiro (RL). .....	93
FIGURA 10	Representação gráfica da suposição de riscos proporcionais para a covariável vida social ativa (VSOA). .....	93
FIGURA 11	Representação gráfica da suposição de riscos proporcionais para a covariável orientação sexual (OS). .....	94
FIGURA 12	Representação gráfica da suposição de riscos proporcionais para a covariável vida sexual ativa (VSA). .....	94
FIGURA 13	Representação gráfica da suposição de riscos proporcionais para a covariável programas sociais (PS). .....	95
FIGURA 14	Representação gráfica da suposição de riscos proporcionais para a covariável escolaridade da gestante (E). .....	95
FIGURA 15	Representações gráficas dos resíduos padronizados de Schoenfeld versus os tempos para as covariáveis consideradas no modelo de Cox estratificado. ....	104
FIGURA 16	Representações gráficas dos $Log(\hat{\Lambda}_{0j}(t))$ versus tempo para as covariáveis OS, RL, VSOA e E. ....	107
FIGURA 17	Representação gráfica da suposição de riscos proporcionais para a covariável OS. ....	108
FIGURA 18	Representação gráfica da suposição de riscos proporcionais para a covariável RL. ....	108
FIGURA 19	Representação gráfica dos resíduos de martingale para o modelo covariáveis tempo-dependente. ....	109
FIGURA 20	Representação gráfica resíduos de deviance para o modelo covariáveis tempo-dependente. ....	109

FIGURA 21	Representações gráficas dos resíduos $\delta\beta$ versus cada covariável no modelo final ajustado para os dados gravidez na adolescência usando modelo de Cox estendido covariável tempo-dependente.....	110
FIGURA 22	Representações gráficas dos resíduos $\delta\beta$ versus cada covariável no modelo final ajustado para os dados gravidez na adolescência usando modelo de Cox estendido covariável tempo-dependente.....	111

## LISTA DE TABELAS

	<b>Página</b>
TABELA 1	Número e proporção de nascimentos cujas mães tinham idade entre 10 e 19 anos, segundo idade da mãe e raça/cor do recém-nascido – Brasil, 2004 (n = 661.290). .... 4
TABELA 2	Prevalência de gravidez antes dos 20 anos entre homens e mulheres, segundo características sociodemográficas ..... 7
TABELA 3	Números de nascidos vivos em função do grau de instrução da mãe no período de 1994 a 2006 no estado de Minas Gerais ..... 8
TABELA 4	Estimativas dos parâmetros do modelo de Cox estratificado para covariável A sem interação com as covariáveis B e C, seus respectivos erros padrões (EP) e valor P. .... 28
TABELA 5	Estimativas dos coeficientes, erros padrões e p-valor para modelo de Cox estratificado para covariável A (estrato A1) sem interação com as covariáveis B e C ..... 28
TABELA 6	Estimativas dos coeficientes, erro padrão (EP) e p-valor para modelo de Cox estratificado para covariável A (estrato A2) sem interação com as covariáveis B e C ..... 28
TABELA 7	Estimativas dos coeficientes, erros padrões e p-valor para modelo de Cox estratificado para covariável A (estrato A1) com interação com as covariáveis B e C ..... 31
TABELA 8	Descrição das covariáveis e das categorias que compõem o banco de dados usado no estudo da gravidez na adolescência na região de Campo Belo, MG ..... 53

TABELA 9	Estimativas dos testes logrank e Wilcoxon e respectivas probabilidades (p-valor) utilizados para testar a igualdade das curvas de sobrevivência para as diversas covariáveis usadas no estudos da gravidez na adolescência na microrregião de Campo Belo - MG.....	58
TABELA 10	Número do modelo e discriminação do modelo usados para obter o modelo paramétrico ajustado aos dados do estudo da gravidez na adolescência na microrregião de Campo Belo- MG .....	59
TABELA 11	Passos usados na seleção de covariáveis, número do modelo, logaritmo da razão de verossimilhança, estatística de teste e p-valor usados para obter o modelo paramétrico ajustado aos dados do estudo da gravidez na adolescência na microrregião de Campo Belo e Região- MG.....	68
TABELA 12	Modelos, Logaritmo da função $L(\theta)$ , resultados da TRV e p-valor usados para obtenção do modelo paramétrico ajustado aos dados do estudo da gravidez na adolescência na microrregião de Campo Belo-MG.....	72
TABELA 13	Estimativas dos parâmetros, erro-padrão, valor de Z e p-valor do modelo de regressão Log- Normal ajustados aos dados do estudo da gravidez na adolescência na microrregião de Campo Belo-MG.....	76
TABELA 14	Número do modelo e descrição do modelo usados para obter o modelo de riscos proporcionais Cox ajustado aos dados do estudo da gravidez na adolescência na microrregião de Campo Belo-MG.....	79
TABELA 15	Passos usados na seleção de covariáveis, número do modelo, modelo, logaritmo da razão de verossimilhança, estatística de teste e p-valor, usados para obter o modelo de riscos proporcionais Cox ajustado aos dados do estudo da gravidez na adolescência na microrregião de Campo Belo-MG .....	83

TABELA 16	Estimativas dos parâmetros ( $\beta$ ), seus respectivos erros-padrão (EP) e p-valor, razão de risco (RR) e intervalo de confiança da razão de risco relativos ao ajuste do modelo de riscos proporcionais de Cox ajustado aos dados da gravidez na adolescência na microrregião de Campo Belo-MG .....	85
TABELA 17	Estimativas dos parâmetros do testes da proporcionalidade dos riscos no modelo de Cox ajustado, valores calculados do teste Qui-quadrado e p-valor .....	96
TABELA 18	Possíveis modelos com estratificação individual versus possíveis modelos com estratificação conjunta para obtenção do modelo ajustado de Cox estratificado para os dados do estudo da gravidez na adolescência na microrregião de Campo Belo - MG.....	98
TABELA 19	Resultados das estimativas das covariáveis, dos erros padrões, da razão de risco e do p- valor de vários modelos ajustados com estratificação individual para os dados do estudo da gravidez na adolescência na microrregião de Campo Belo-MG.....	99
TABELA 20	Continuação dos resultados das estimativas das covariáveis, dos erros-padrão, da razão de risco e do p-valor de vários modelos ajustados com estratificação conjunta para os dados do estudo da gravidez na adolescência em Campo Belo e Região - MG.....	101
TABELA 21	Resultado das estimativas dos parâmetros, dos erros-padrão (EP), do p-valor, da razão de risco(RR) e do intervalo de confiança da razão de risco do ajuste do modelo de Cox estratificado com estratificação conjunta (E+PS+VSA+Strata(OS+RL)+VSOA +Rel+Ren) para explicar os dados da gravidez na adolescência na microrregião de Campo Belo -MG.....	103

TABELA 22	Estimativas dos parâmetros, valores do teste Qui-quadrado e p-valor para aplicação do teste da proporcionalidade dos riscos no modelo de Cox ajustado.....	106
TABELA 23	Resultado das estimativas dos parâmetros, do erro padrão (EP), do p-valor, da razão de risco (RR) e do intervalo de confiança da razão de risco do ajuste do modelo de Cox com variável tempo-dependente para explicar os dados da gravidez na adolescência na microrregião de Campo Belo-MG.....	113
TABELA 24	Estimativas dos coeficientes, Erro-Padrão e Exponencial das estimativas dos coeficientes para os quatros dos modelos ajustados.....	114
TABELA 25	Tempo mediano(TM) e razão de risco (RR) das estimativas dos coeficientes dos modelos ajustados.....	114
TABELA 26	Intervalos de confiança do tempo mediano das estimativas dos coeficientes do modelo paramétrico Log-Normal e da razão de risco dos modelos de Cox.....	115

## RESUMO

SANTOS, Aladir Horacio. **Modelagem e análise de fatores influentes na gravidez na adolescência na microrregião de Campo Belo-MG**. 2009. 124p. Tese (Doutorado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, MG.\*

A adolescência é um período de transição e atribuir-lhe uma definição que contemple tanto seus aspectos fisiológicos quanto psicossociais é uma tarefa difícil. Segundo a Organização Mundial da Saúde, “é uma etapa evolutiva caracterizada pelo desenvolvimento biopsicossocial, delimitada como a faixa etária entre dez e dezenove anos”. Um aspecto dessa faixa etária que vem sendo fonte de grande preocupação por parte dos órgãos de saúde pública, principalmente nos países subdesenvolvidos, é a gravidez na adolescência. É nesse contexto e perspectiva que se encontra a proposta deste estudo, que realizou uma pesquisa com todas as mulheres que deram à luz na maternidade da Santa Casa de Misericórdia da cidade de Campo Belo no período de 2000 a 2006. A pesquisa envolveu a aplicação de um questionário para verificar quais fatores -- entre eles renda, escolaridade, métodos contraceptivos --, mais contribuem para a gravidez na adolescência. A técnica estatística usada foi a análise de sobrevivência. Os dados foram analisados de forma não-paramétrica e semi-paramétrica (modelos de Cox: riscos proporcionais, estratificados e variável tempo dependente), além de se basear também em modelos estatísticos paramétricos. Concluiu-se que todas as suspeitas relacionadas por vários pesquisadores sobre os fatores que contribuem pelo agravamento da ocorrência da gravidez na adolescência relacionados nesse trabalho foram confirmadas.

Palavras - Chave: Gravidez na Adolescência; Saúde Pública; Análise de Sobrevivência; Modelos de Cox.

---

\* Comitê Orientador: Prof<sup>o</sup> Dr<sup>o</sup> Augusto Ramalho de Moraes – UFLA (Orientador) e Prof<sup>o</sup> Dr<sup>o</sup> Mario Javier Ferrua Vivanco – UFLA (Co-orientador).



## ABSTRACT

SANTOS, Aladir Horacio. **Modeling and analysis of factors influencing adolescent pregnancy in the micro region of Campo Belo-MG**. 2009. 124p. Thesis (Doctorate in Statistics and Agricultural Experimentation) – Federal University of Lavras, Lavras, Minas Gerais, Brazil. \*

The adolescence is a transitional period that occurs between childhood and adulthood which multifaceted nature cannot be measured, at the same time, by any definition alone. According to the World Health Organization, it is “an evolutionary stage characterized by a rapid biological (i.e. pubertal), social and psychological development, which occurs in between 10 and 19 years of age”. Given these complex and delicate characteristics, one of the issues that have most concerned public health authorities, particularly in underdeveloped countries, is the adolescent pregnancy. This study aims and scope sought to fit in this context by carrying out a research in the maternity of the Santa Casa de Misericórdia located in the city of Campo Belo. A survey was applied to all women who gave birth at this maternity from 2000 to 2006. The survey questions were designed in order to investigate which factors -- such as, for instance, the level of schooling, income, birth control methods --; most contribute to the adolescent pregnancy. The statistical technique used was the Survival analysis. The data were analyzed through non-parametric and semi-parametric tests (Cox proportional hazards model, stratified, and time is the response variable) and through parametric statistical models. The conclusion confirmed other research findings on all factors already reported which contribute most to the adolescent pregnancy.

Keywords: pregnancy; adolescence; survival analysis; Cox proportional hazards model

---

\* Guidance Committee: Prof. DSc. Augusto Ramalho de Moraes – UFLA (Adviser) e Prof. DSc. Mario Javier Ferrua Vivanco – UFLA (co-Adviser).

## 1 INTRODUÇÃO

A adolescência é um período de transição e atribuir-lhe uma definição que contemple tanto seus aspectos fisiológicos quanto psicossociais é uma tarefa difícil. Segundo a Organização Mundial da Saúde (*World Health Organization - WHO, 1975*), “é uma etapa evolutiva caracterizada pelo desenvolvimento biopsicossocial, delimitada como a faixa etária entre dez e dezenove anos”, marcada por importantes transformações anatômicas tais como alterações na composição corporal - estirão de crescimento e fisiológicas: maturação sexual, gonadal, esquelética, cardiovascular, respiratória, mentais ou psicológicas e sociais.

Um aspecto correlacionado às mudanças ocorridas nesta faixa etária e que vem sendo fonte de grande preocupação por parte dos órgãos de saúde pública -- principalmente nos países subdesenvolvidos -- é a gravidez na adolescência.

De acordo com Brasil 2006, durante o ano de 2004, 21,9% dos recém-nascidos no país tinham mães com idade entre 10 e 19 anos, e 0,9% tinham mães com idade entre 10 e 14 anos, constatando um aumento relativo da fecundidade de adolescentes em relação à de mulheres mais velhas.

Entre os vários fatores considerados influentes na gravidez precoce, destacam-se o baixo poder aquisitivo, o baixo nível de escolaridade materna, adolescentes cujas mães também engravidaram na adolescência e iniciação sexual precoce. São citados também outros fatores de risco, tais como adolescentes órfãs de pai ou mãe, história familiar de gravidez na adolescência, gravidez vista como valor social positivo e comportamentos de risco (como uso de álcool e/ou outras drogas). Como fatores de proteção, foram mencionados níveis escolares e socioeconômicos altos, famílias com pai e mãe, engajamento em atividades religiosas, dentre outros (Elfenbein & Felice, 2003).

Poucos são os trabalhos de cunho estatístico que relacionam os fatores de risco com a ocorrência da gravidez na adolescência. Entre as técnicas estatísticas, se destaca a análise de sobrevivência, cujo propósito é analisar dados de tempo de vida, sendo que a variável resposta é o tempo até a ocorrência do evento de interesse, no caso presente, a ocorrência da gravidez.

É neste contexto e perspectiva que se encontra a proposta deste estudo, ao realizar uma pesquisa junto à maternidade da Santa Casa de Misericórdia da cidade de Campo Belo - Minas Gerais- e cidades vizinhas, com o intuito de identificar os possíveis fatores determinantes da gravidez na adolescência nesta região.

De modo específico, um dos objetivos foi o de fazer uma modelagem estatística da ocorrência da gravidez na adolescência, procurando identificar seus fatores mais relevantes, visando a auxiliar a tomada de decisão das políticas públicas, bem como o seu planejamento.

## 2 REFERENCIAL TEÓRICO

### 2.1 Gravidez na adolescência

No Brasil, tem havido um aumento da população adolescente nas últimas décadas. Uma das razões para esse aumento se deve, principalmente, a uma atenção maior dispensada à população infantil, abrangendo ações públicas de grande alcance tais como campanhas de vacinação e combate à desnutrição, entre outras, assim reduzindo o índice de mortalidade infantil (Yazlle, 2006).

De acordo com os dados do DATASUS (Brasil, 2009) de 1994 a 1999 ocorreram 17.772.853 partos, dos quais 145.487 eram de adolescentes na faixa etária de 10 a 14 anos e 3.800.685 eram adolescentes na faixa etária de 15 a 19 anos. Já de 2000 a 2006, período em que foi realizada a pesquisa, ocorreram 21.426.460 nascidos vivos, dos quais 192.445 eram de adolescentes na faixa etária de 10 a 14 anos e 4.604.431 eram de adolescentes na faixa etária de 15 a 19 anos. Nota-se que a porcentagem de nascidos vivos aumentou 20,5% no período. No entanto, o percentual de nascidos vivos, cujas mães eram adolescentes na faixa etária de 10 a 14 anos, aumentou 32,3%, enquanto que o percentual de nascidos vivos de mães adolescentes na faixa etária de 15 a 19 anos aumentou 21,1%.

De acordo com Brasil (2006), Tabela1, durante o ano de 2004, o Sinasc captou 3.026.548 nascimentos no país; 661.290 (21,9%) recém-nascidos tinham mães com idade entre 10 e 19 anos e 26.276 (0,9%) tinham mães com idade entre 10 e 14 anos. Assim, dentre os nascimentos cujas mães tinham idade até 19 anos, 4% foram de mães que tinham até 14 anos. A maior parte dos nascimentos cujas mães eram adolescentes foi de bebês de raça/cor negra (329.616 bebês, o que corresponde a 54,4%) (BRASIL, 2006).

TABELA 1 Número e proporção de nascimentos cujas mães tinham idade entre 10 e 19 anos, segundo idade da mãe e raça/cor do recém-nascido – Brasil (n = 661.290)

<i>Idade (em anos)</i>	<i>Nascimento</i>	<i>%</i>
10	11	0,002
11	115	0,02
12	691	0,1
13	4.720	0,7
14	20.739	3,1
15	55.202	8,3
16	100.120	15,1
17	136.283	20,6
18	166.127	25,1
19	177.252	26,6
<b>Raça/cor do recém-nascido</b>		
Branca	269.359	44,5
Preta	13.631	2,2
Parda	315.985	52,1
Indígena	4.403	0,7
Amarela	2.603	0,4

Fonte: Sinasc/SVS/MS

Esse aumento de gravidez precoce tem sido considerado por muitos como um problema social, pois geralmente as adolescentes apresentam incapacidade fisiológica para gestar e incapacidade psicológica para criar (Simões et al., 2003; Menezes & Domingues, 2004; Michelazzo et al., 2004; Esteves & Menandro, 2005). Além disso, muitos outros problemas de saúde -- tanto da mãe quanto da criança -- podem surgir, assim como problemas de ordem social.

Ao realizar uma descrição sobre a gravidez na adolescência ao longo do tempo nos Estados Unidos, Elfenbein & Felice (2003) defendem que, embora haja taxas significativamente discrepantes de gravidezes e nascimentos entre adolescentes de diferentes grupos raciais, há similaridades entre eles. Baixo poder aquisitivo, baixo nível de escolaridade materna, adolescentes cujas mães também engravidaram na adolescência e iniciação sexual precoce, foram alguns

dos fatores de semelhança (e de risco) citados. Outros fatores de risco foram adolescentes órfãs de pai ou de mãe, história familiar de gravidez na adolescência, gravidez vista como valor social positivo e outros comportamentos de risco (como uso de álcool e outras drogas). Como fatores de proteção, foram mencionados níveis escolares e socioeconômicos altos, famílias com pai e mãe, engajamento em atividades religiosas, dentre outros (Elfenbein & Felice, 2003).

Quanto à evolução da gestação, diversos autores citam a gravidez na adolescência como um fator de predisposição ao risco devido à incidência maior de complicações no parto e puerpério, uma vez que a gravidez ocorre em um organismo que ainda está em amadurecimento físico e emocional (Santos et al., 2008). Segundo Yazlle (2006, p.6).

Existem referências a maior incidência de anemia materna, doença hipertensiva específica da gravidez, desproporção céfalo-pélvica, infecção urinária, prematuridade, placenta prévia, baixo peso ao nascer, sofrimento fetal agudo intra-parto, complicações no parto (lesões no canal de parto e hemorragias) e puerpério (endometrite, infecções, deiscência de incisões, dificuldade para amamentar, entre outros).

Assim, segundo Rodrigues et al. (1993) as complicações da gravidez, parto e puerpério constituem a 10ª causa de óbitos entre adolescentes brasileiras.

Santos et al. (2008) corroboram as complicações acima mencionadas, explicitando que estudos demonstram frequência maior de recém-nascidos de BPN (baixo peso ao nascer), sobretudo nas adolescentes entre 10 e 15 anos, provavelmente pelo baixo peso materno anterior à gestação, ganho ponderal insuficiente, conflitos familiares e existenciais que retardam a procura pela assistência pré-natal, maior incidência de anemia e infecções e incompleto desenvolvimento nos órgãos reprodutivos, que podem acarretar insuficiência placentária, prejudicando as trocas materno-fetais – além do fato de que os efeitos de uma gravidez na adolescência antes de o próprio desenvolvimento

materno se completar poderiam estar associados com o risco aumentado de parto pré-termo e BPN. O BPN é o fator mais importante, associado à mortalidade e morbidade perinatais, quando se avalia o desfecho da gravidez, além de ser o indicador isolado mais importante de morbimortalidade infantil. Estudos demonstram que a incidência de BPN é mais que o dobro em adolescentes em relação às mulheres adultas.

Segundo Souza et al. (1999), ainda dentro das complicações médico-obstétricas, a que mais se associa a danos físicos e psicológicos é o aborto espontâneo ou provocado (Souza, 1999). Além disso, de acordo com Del Ciampo citado por Santos et al. (2008), há danos ainda na qualidade de vida e na perspectiva de futuro dessas adolescentes, uma vez que as mesmas geralmente encontram dificuldades para retomar aos estudos e/ou dificuldades de reinserção no mercado de trabalho (WHO, 1975).

A educação está entre os fatores de proteção, pois adolescentes com maior escolaridade são menos propensas à gravidez precoce. Isso pode ser confirmado na Tabela 2, citada por Aquino et al. (2003), Baraldi et al. (2007) e Ximenes Neto et al. (2007). Note que, à medida em que a escolaridade aumenta, diminui a prevalência de gravidez antes dos 20 anos. Pode-se chegar a essa mesma conclusão observando-se os dados do DATASUS, referentes ao Estado de Minas Gerais no período compreendido entre 1994 e 2006 (Tabela 3). Se considerarmos o período de 1999 a 2006, que é o período em que se tem maior quantidade de dados completos, notamos que a quantidade de mulheres com instrução até sete anos é inferior à de mulheres com instrução superior a 8 anos. Esse fato ocorreu até o ano de 2003, a partir do qual ocorreu a situação inversa.

TABELA 2 Prevalência de gravidez antes dos 20 anos entre homens e mulheres, segundo características sociodemográficas

Características sociodemográficas	Mulheres			Homens		
	n	%	p-valor	n	%	p-valor
<b>Renda familiar per capita*</b>						
Muito baixo (até R\$ 90,00)	381	50,4	0,0000	267	33,9	0,0000
Baixo (de R\$91,00 até R\$180,00)	487	36,8		396	27,6	
Intermediário (de R\$181,00 a R\$540,00)	833	24,4		800	16,8	
Alta (acima de R\$540,00)	734	7,8		712	11,2	
<b>Escolaridade</b>						
1º grau incompleto	437	59,6	0,0000	496	34,4	0,0000
1º grau completo	575	36,8		565	32,2	
2º grau completo	796	17,0		635	13,1	
Superior incompleto/completo	597	4,6		447	6,5	
<b>Raça/Cor</b>						
Branca	1.165	20,5	0,0000	1.099	14,6	0,0000
Parda	528	31,1		442	17,3	
Preta	550	37,4		465	31,1	
Indígena	133	30,8		146	24,4	
<b>Religião Atual</b>						
Católica	839	21,5	0,0030	539	20,3	0,2997
Pentecostal	246	30,0		178	22,4	
Outra	451	28,0		259	13,1	
Nenhuma	882	33,6		1.145	20,5	
<b>Trabalho</b>						
Sim	1.083	24,0	0,0066	1.317	21,6	0,2211
Não	1.352	30,6		857	17,1	

\*Categoria definida com base no salário mínimo nacional na época (R\$180,00).



TABELA 3 Números de nascidos vivos em função do grau de instrução da mãe no período de 1994 a 2006 no estado de Minas Gerais

Instrução da mãe	Ano						
	1994	1995	1996	1997	1998	1999	
Nenhuma	1552	3809	6084	6761	12548	65	
1 a 3 anos	-	-	0	0	0	241	
4 a 7 anos	-	-	0	0	0	315	
8 a 11 anos	-	-	0	0	0	226	
12 anos e mais	-	-	9211	9550	12377	136	
9 a 11 anos	-	-	26184	28552	43169	52	
1º grau incompleto	33208	67385	96402	110725	172475	158	
1º grau completo	8065	15235	24256	25448	39551	55	
2º grau	9540	17270	-	-	-	-	
Superior	4028	6862	-	-	-	-	
Ignorado	5751	12235	13806	10978	15244	306503	
<b>Total</b>	<b>62144</b>	<b>122796</b>	<b>175943</b>	<b>192014</b>	<b>295364</b>	<b>307751</b>	

Instrução da mãe	Ano						
	2000	2001	2002	2003	2004	2005	2006
Nenhuma	7638	7366	5780	5458	4421	3841	3250
1 a 3 anos	43628	40654	34756	31556	28041	25026	20792
4 a 7 anos	123862	119354	113121	108509	102555	96467	86823
8 a 11 anos	77625	85020	86669	95119	101346	110317	110421
12 anos e mais	31022	32432	32506	33230	33823	34735	38691
9 a 11 anos	99	0	-	-	-	-	-
1º grau incompleto	90	36	-	-	-	-	-
1º grau completo	92	5	-	-	-	-	-
2º grau	-	-	-	-	-	-	-
Superior	-	-	-	-	-	-	-
Ignorado	16902	13671	11726	11032	7505	7082	6166
<b>Total</b>	<b>300958</b>	<b>298538</b>	<b>284558</b>	<b>284904</b>	<b>277691</b>	<b>277468</b>	<b>266143</b>

Fonte: MS/SVS/DASIS - Sistema de Informações sobre Nascidos Vivos – SINASC.

O mesmo fenômeno também foi relatado por Paraguassú et al. (2005) ao observarem um alto índice de adolescentes grávidas entre 17 e 19 anos que ainda estavam cursando o ensino fundamental.

A escolaridade do pai e/ou da mãe é outro fator que contribui para o problema em questão. Caputo & Bordin (2008) afirmam que estudos apontam

associações significativas entre gravidez na adolescência e a escolaridade dos pais.

Citações anteriores mencionam que adolescentes com maior escolaridade são menos propensas à gravidez, o que mostra que a escola é uma das instituições sociais com capacidade de intervir no problema em questão. A educação sexual faz parte dos Parâmetros Curriculares Nacionais (PCNs) que orientam o trabalho nas escolas do País desde 1997. O texto não fala em obrigatoriedade de uma disciplina específica para o tema, apenas sugere que ele faça parte do projeto pedagógico da escola e seja trabalhado em todos os conteúdos. Não basta que os projetos pedagógicos envolvam somente as crianças e os adolescentes, mas devem estender-se aos pais, que, em sua maioria, estão despreparados para tratar desta questão com os filhos. Às vezes, a adolescente até mesmo deseja lhes contar suas experiências, mas muitos não querem ouvir ou fantasiam ter uma eterna criança dentro de casa.

Outro importante fator de proteção é a igreja. Vários autores têm mostrado que a religião tem tido uma participação significativa para a diminuição da gravidez na adolescência. Entre eles, Aquino et al. (2003) comenta que mulheres da religião católica relataram apresentar menor experiência sexual antes dos vinte anos, como mostra a Figura 1. Já Carvalho (1999) e Guimarães (2001) relataram que as novas religiões evangélicas que têm florescido apresentam maior rigidez no que diz respeito à prática sexual.

Dessa maneira, fica claro o importante papel orientador que as instituições religiosas exercem, já que são poderosos meios de comunicação e formadores de opinião.

Aquino et al. (2003) relatam que as adolescentes que dialogam com suas mães sobre menstruação antes da menarca ou cujos pais (ou a mãe ou pai) foi(ram) incluído(s) entre as primeiras fontes de informação sobre gravidez e métodos anticoncepcionais engravidam menos na adolescência, o que mostra o

papel de destaque da família como fator de proteção, o que pode ser observado na Tabela 2.

No entanto, entre outros fatores, a falta de compreensão, diálogo, afeto pode levar a adolescente a tomar medidas cujas consequências são imensuráveis. Monteiro et al. (2007) comentam que, pelo fato de as adolescentes não apresentarem suporte físico e nem emocional, a gravidez precoce e não planejada pode levar a situações de conflito com a família, o que pode chegar a consequências drásticas, como por exemplo, o suicídio.

Já Young et al. (1991) argumentam que a probabilidade de engravidar durante a adolescência é maior para filhas de pais separados ou solteiros.

Do exposto acima, fica clara a importância da família para a diminuição da gravidez na adolescência.

De acordo com Strasburger (1999), entre os fatores de risco, pode-se citar a mídia, pois muitos são os filmes, novelas, além de outras formas de entretenimentos que lidam com o sexo na adolescência de maneira e em situações inadequadas, onde a relação sexual e a contracepção têm um tratamento distanciado, isto é, foca-se o ato sexual, mas não suas implicações, o que pode favorecer a promiscuidade.

As cenas de sexo explícito estão cada vez mais frequentes em novelas, filmes, e demais programas veiculados pela mídia, porém nem sempre com a mesma frequência as questões dos métodos contraceptivos e das consequências de o não-uso dos mesmos pode acarretar (como a gravidez indesejada) são mostradas e/ou debatidas. Assim, não é surpreendente que algumas adolescentes imaturas, por encararem as situações da dramaturgia como reproduções fidedignas da vida real, incorram em uma postura de imitação, o que as conduz a um comportamento de risco.

Berlofi et al. (2006), em um trabalho sobre os efeitos de Programas de Planejamento Familiar, relatam que o início da atividade sexual está ocorrendo

de forma precoce, o que pode aumentar o número de adolescentes grávidas e apontam como possíveis responsáveis o estilo de vida, fatores do meio ambiente e a mídia.

Outro fator de risco é o uso de drogas ilícitas. Caputo & Bardin (2008) relatam que, além da baixa escolaridade paterna, ausência de informação sobre sexualidade e fertilidade, o uso de drogas ilícitas por familiar residente no domicílio também está associado à gravidez na adolescência.

A cor também tem se mostrado um fator de risco. Aquino et al. (2003) mostram que o percentual de adolescentes grávidas é menor entre as adolescentes brancas do que entre outras pertencentes a grupos de cor diferentes, conforme ilustrado na Tabela 2. Sabroza et al. (2004) observaram um predomínio de puérperas de cor parda e negra em relação às de cor branca. Sabroza & Henshaw (2007) citados por Chalem et al. (2007) comentam que os maiores índices de gestações na adolescência recaem preferencialmente sobre a parcela negra da população. Talvez uma explicação para tal afirmação sejam as condições socioeconômicas precárias em que vive essa parte da população.

Os métodos anticoncepcionais (MAC) têm-se revelado um forte fator de risco. Em muitos trabalhos, encontram-se citações de que aproximadamente metade das adolescentes engravidam de forma indesejada. Entre diversas razões que contribuem para esse fato, cita-se o não uso ou desconhecimento dos MAC (Silva et al., 2007).

Em estudo sobre conhecimento, atitude e prática sobre MAC entre as adolescentes grávidas, Belo & Silva (2004) relatam que é significativa a quantidade de informação para as adolescentes. No entanto, os autores concluem que a disponibilidade do conhecimento, de mais serviços e dos próprios MAC não foram suficientes para permitirem uma mudança de postura das adolescentes frente à prática do uso eficiente destes.

É importante notar também que, mesmo tendo conhecimento dos MAC, as adolescentes não o utilizam apropriadamente, pois, além da dificuldade de acesso, têm receio dos efeitos colaterais, acreditam que são imunes à gravidez, não conhecem o próprio corpo, não conseguem colocar o assunto em discussão na família e tampouco recebem qualquer orientação na escola, uma vez que ainda persiste o mito de que falar sobre sexo estimula sua prática (Cabral, 2003).

Outro fator que pode contribuir para o problema é a falta de lazer. Em muitas cidades, a única opção de lazer para os jovens é beber nos barezinhos, ir a festas, a exposições, ou a outras atividades locais cujas opções são, muitas vezes, limitadas, pouco criativas. Oferecer alternativas de lazer e possibilidades de esporte que resgatem o lado lúdico e recreativo é também uma forma de prevenção.

A gravidez ainda representa para muitas meninas brasileiras uma das únicas alternativas de elevar seu status social. A conclusão é da coordenadora do Programa de Saúde do Adolescente da Secretaria de Estado de Saúde do Rio de Janeiro, a psicóloga Tânia Catharino. Durante o 13º Encontro da Associação Brasileira de Estudos Populacionais (ABEP), ela apresentou um estudo qualitativo com adolescentes que deram à luz no município do Rio entre 1999 e 2001, quando tinham de 10 a 14 anos de idade. Catharino alerta que a informação e o acesso a métodos anticoncepcionais não são suficientes para combater a gravidez precoce. Dependendo do contexto social e apesar do ônus que acarretam, essa gravidez traz ganhos secundários para a adolescente. Segundo Miguel<sup>3</sup> (2005) “após a adolescente ficar grávida, ela passa a receber atenção especial, que inclui ajuda financeira, material e afetiva”.

---

<sup>3</sup> Comunicação pessoal: MIGUEL, Denise Freire Assumpção Campo Belo: Secretaria de Saúde, 2005.

## **2.2 Análise de sobrevivência**

O estudo da análise de sobrevivência centra-se num conjunto de unidades (indivíduos) que são observados até a ocorrência de algum evento de interesse, por exemplo, a falha das unidades. De acordo com Rocha (1995), essas unidades dão origem aos dados de sobrevivência que são formados essencialmente pelos tempos de vida dessas unidades até a ocorrência de algum evento de interesse. Por exemplo, os tempos decorridos entre o diagnóstico de uma determinada doença em pacientes até a sua morte ou até a sua cura. Estes tempos de vida são algumas vezes chamados de tempo de sobrevivência, podendo também ser referidos como tempos de falha ou tempos de vida. Existem algumas situações em que esse evento não chega a ocorrer para algumas unidades durante o período de observação, o que é caracterizado como censura. As causas associadas à ocorrência de censura são bastante variadas e, por conseguinte, existem vários tipos de censura ou observações incompletas.

### **2.2.1 Censura**

Na análise de sobrevivência, uma característica importante dos dados em estudo é a observação incompleta sobre o tempo de falha do indivíduo, comumente chamado de censura. Colosimo & Giolo (2006) argumentam que mesmo sendo incompletas, as observações censuradas fornecem informações sobre o tempo de vida e que a omissão das censuras no cálculo das estatísticas de interesse certamente acarretará resultados viciados. Portanto, mesmo censurados, todos os resultados provenientes de estudo de sobrevivência devem ser usados na análise estatística.

Rocha (1995) classifica a censura da seguinte maneira: i) censura à direita: é aquela em que o tempo entre o início e o evento é maior do que o tempo observado e é o mais encontrado em estudo de tempo de vida. Nesse caso, aproveita-se a informação sem que o evento tenha ocorrido. Desprezar essa

informação faria com que o risco fosse superestimado. ii) censura à esquerda: é o tipo de censura em que o tempo de ocorrência do evento de interesse aconteceu em um tempo inferior ao tempo registrado, isto é, não se conhece o momento da ocorrência do evento, mas sabe-se que ele ocorreu antes de um tempo determinado. iii) censura intervalar: é aquela em que o tempo de ocorrência do evento de interesse ocorreu em um determinado intervalo de tempo. A censura intervalar pode ser considerada como uma combinação das censuras à direita e à esquerda. iv) tipo I – aquela em que o estudo é encerrado após um intervalo de tempo preestabelecido. v) tipo II – aquela em que o estudo é encerrado após ter ocorrido a falha em um número definido de indivíduos; vi) aleatória – ocorre quando o indivíduo é retirado do estudo por causas adversas ao experimento, ou por razões que fogem ao controle do pesquisador.

Outro tipo de dado incompleto, denominado dado truncado, ocorre quando indivíduos são excluídos do estudo por motivo relacionado à ocorrência do evento. Em estudos com truncamento, somente indivíduos que experimentam algum evento são incluídos. O truncamento de acordo com Rocha (1995) pode ser: à esquerda quando a perda da informação está relacionada a indivíduos que foram excluídos porque já tinham experimentado o evento antes do início do estudo e não podiam ser observados e à direita quando o critério de seleção dos indivíduos inclui somente os que sofreram o evento.

### **2.3 Inferência não-paramétrica**

Se em um conjunto de dados não há indivíduos censurados, a função de sobrevivência pode ser estimada a partir dos dados observados como a proporção dos indivíduos que sobreviveram além do tempo. No entanto, havendo dados censurados, como estimar a função de sobrevivência? Uma das maneiras é utilizar um estimador não-paramétrico para a função de sobrevivência  $S(t)$ . Entre alguns possíveis estimadores de  $S(t)$ , o mais citado na

literatura é o estimador de Kaplan & Méier (1958), também conhecido como estimador Produto-Limite. Segundo Carvalho et al. (2005), este estimador utiliza os conceitos de independência condicional para desdobrar a condição de sobreviver até o tempo  $t$  em uma sequência de elementos independentes que caracterizam a sobrevivência em cada intervalo de tempo anterior a  $t$  e cuja probabilidade é condicional aos que estão em risco em cada período.

O estimador de Kaplan & Méier (1958) é definido por:

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

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

## 2.4 Modelos de regressão

### 2.4.1 Variáveis explanatórias

Como já mencionado o objetivo fundamental da Análise de Sobrevivência é a análise do tempo de vida de cada indivíduo. No entanto, é plausível admitir que a sua sobrevivência possa ser afetada por diversos fatores, designados por fatores de risco, tais como tratamentos, propriedades intrínsecas do indivíduo ou variáveis exógenas. Assim, a partir da década de 70, foi dada uma ênfase particular ao estudo da associação entre o tempo de vida e variáveis designadas por explanatórias ou covariáveis, representando os fatores acima mencionados. Sempre que tal for possível, os valores individuais destas



variáveis devem ser registrados, visto que fornecem informação acerca da heterogeneidade existente na população.

Segundo Rocha (1995), as covariáveis podem ser dependentes do tempo ou constantes e os seus valores são medidos ao longo do período de observação ou uma única vez. Existe, no entanto, outra classificação mais detalhada no que diz respeito à dependência do tempo. Assim, as covariáveis podem ser classificadas em duas categorias principais; covariáveis externas ou internas.

i) Diz-se que uma covariável é externa se esta não está diretamente relacionada ao o mecanismo que regula a falha dos indivíduos. Pode ainda ser classificada como:

ii) Fixa: quando o valor da covariável é medido no início do estudo e permanece constante durante todo o período em que o indivíduo se encontra em observação; b) definida: se a sua trajetória, embora não sendo constante, é determinada a priori para cada indivíduo. É o caso de um fator sob controle do experimentador, que o faz variar ao longo do estudo de forma pré-determinada; c) ancilária: quando é o resultado de um processo que é exterior ao indivíduo. Por exemplo, uma covariável que representa o nível de poluição atmosférica num estudo da ocorrência de ataques de asma.

ii) Diz-se que uma covariável é interna se esta resulta de uma medida feita sobre um indivíduo ao longo do tempo. Logo, é observada apenas enquanto o indivíduo está vivo e não censurado, fornecendo deste modo informação sobre o seu tempo de vida. Numa experiência clínica, como exemplo deste tipo de covariáveis, é uma determinada medida da condição geral do paciente, feita a intervalos regulares.

#### **2.4.2 Formulação de um modelo de regressão**

Quando se pretende modelar o tempo de vida numa população homogênea, é habitual utilizar determinadas distribuições contínuas univariadas

que, por apresentarem certas propriedades, são particularmente adequadas num grande número de situações práticas. É o caso das distribuições Exponencial, Gama, Weibull, Gompertz, Log-Normal, Log-Logística e outras, conforme referido em Kalbfleisch & Prentice (1980), Lawless (1982) e outros.

No entanto, a situação mais frequente entre os indivíduos no que diz respeito a possíveis fatores de risco é a existência de heterogeneidade. Na análise estatística, uma forma adequada de incorporar esses fatores que, se supõem, afetam o tempo de vida, consiste na utilização de um modelo de regressão, em que o tempo de vida é a variável dependente e as covariáveis atuam como variáveis independentes. É necessário então especificar um modelo para a distribuição do tempo de vida  $T$  dado o vetor  $x = (x_1, x_2, \dots, x_p)'$  de variáveis explanatórias associado a determinado indivíduo, o que pode ser feito usando alguma família paramétrica de distribuições ou recorrendo a uma abordagem não paramétrica.

De modo geral, é conveniente definir o vetor de covariáveis de modo que  $x=0$  corresponde a algum conjunto de condições padrão. No que se segue  $T$  é uma variável aleatória (v.a.) contínua e  $\varphi(x)$  e  $\alpha(x)$  designam funções que relacionam o vetor  $z$  com a sobrevivência do indivíduo de maneira que, uma função crescente corresponde sempre a tempo de vida decrescente.

A maior parte dos modelos de regressão utilizados em análise de sobrevivência pertencem a duas vastas classes: modelos com funções de riscos proporcionais (PH) e modelos de tempos de vida acelerados (AFT). Segundo Rocha (1995) o que difere um do outro é que a hipótese da AFT é aplicada para a comparação do tempo de sobrevivência, enquanto que a hipótese PH é aplicada para comparar o risco.

## 2.5 Modelos de regressão

O objetivo de um modelo de regressão é estimar o efeito de covariáveis, também chamadas variáveis independentes ou preditoras, sobre uma variável resposta. De acordo com a distribuição que acreditamos ter a variável resposta e, de acordo com a função escolhida para relacioná-la às covariáveis, identifica-se o modelo de regressão como: linear, poisson, logístico, etc.

Em muitas situações, o tempo de vida pode ser influenciado por uma ou mais variáveis regressoras ( $x$ ) que podem ser qualitativas ou quantitativas. Colosimo & Giolo (2006) dizem que as técnicas não-paramétricas não permitem a inclusão de covariáveis na análise, pois estas não envolvem nenhuma estrutura paramétrica e que uma maneira eficiente de se determinar o relacionamento entre  $T$  e  $x$  é por meio de modelos de regressão apropriados para os dados censurados, nos quais a dependência do tempo de vida com as variáveis é explicitamente examinada.

## 2.6 Modelos de riscos proporcionais de Cox

O modelo de regressão de Cox -- proposto por Cox (1972) -- é um dos modelos mais utilizados em análise de sobrevivência. Ele permite a inserção de covariáveis de interesse na análise dos tempos de vida até a ocorrência de um evento. Cox (1975) propõe modelar dados de sobrevivência, na presença de covariáveis, por meio da função de risco.

O modelo de Cox assume que os tempos  $t_i$ ,  $i = 1, 2, 3, \dots, n$ , são independentes e que o risco do indivíduo  $i$  é dado por:

$$\alpha_i(t_i | x_i) = \alpha_o(t) \exp(\beta' x_i) \quad (2)$$

Em que  $\alpha_o$  é conhecido como função de risco base, sendo o risco de um indivíduo sem a presença de covariáveis;  $\beta$  é o vetor de coeficientes de regressão, com dimensões  $p \times 1$ , e  $x_i$ , é o vetor de covariáveis observadas, com dimensões  $p \times 1$ , para o indivíduo  $i$ .

O modelo definido em (2), além de ser conhecido como modelo de regressão de Cox, é também conhecido como modelo semi-paramétrico por assumir que as covariáveis atuam multiplicativamente no risco pela relação  $g(x, \beta) = \exp(\beta'x_i)$  e por considerar  $\alpha_o(t)$  arbitrária, isto é, não há forma paramétrica para  $\alpha_o$ . Esse modelo é denominado modelo de riscos proporcionais devido ao fato de a razão entre as funções de risco de dois indivíduos quaisquer não depender do tempo,

$$\frac{\alpha(t | x_i)}{\alpha(t | x_l)} = \frac{\alpha_o(t) \exp(\beta'x_i)}{\alpha_o(t) \exp(\beta'x_l)} = \exp\{\beta'(x_i - x_l)\}$$

Sendo  $(i, l = 1, 2, 3, \dots, n \text{ e } i \neq l)$

A presença do termo não paramétrico  $\alpha_o$  absorve o termo constante do modelo, por isso a constante  $\beta_o$  não parece no componente paramétrico  $\exp(\beta'x_i)$ .

A suposição básica para o uso do modelo de riscos proporcionais de Cox é, portanto, que as taxas de falha sejam proporcionais. Assim, se um indivíduo apresentar, no início do estudo, um risco de falha igual a duas vezes o risco de um outro indivíduo, essa taxa será a mesma para qualquer tempo  $t$  do período de acompanhamento.

O modelo de Cox é muito usado em estudos clínicos por apresentar uma flexibilidade devido à presença do componente não-paramétrico.

De acordo com Lawless (1982), a função de taxa de falha base acumulada é representada por

$$H_o(t) = \int_0^t \alpha_o(u) du \quad (3)$$

a função de sobrevivência é dada por

$$S(t | x) = \exp\{-H_o(t) \exp(\beta'x)\} = (S_o(t))^{\exp(\beta'x)} \quad (4)$$

e a função de risco base dada por

$$S_o(t) = \exp\{-H_o(t)\} \quad (5)$$

## 2.7 Estimação dos parâmetros no modelo de Cox

Uma característica do modelo de Cox é a estimação dos coeficientes  $\beta$  por meio de observações amostrais. Entretanto, o método de estimação por máxima verossimilhança é inviabilizado devido à presença do componente não-paramétrico. Para contornar tal situação, isto é, para a estimação do vetor de parâmetros desconhecidos  $\beta$ , Cox (1975) propôs a utilização da função de verossimilhança parcial,  $L(\beta)$ , que consiste em tomar uma amostra de  $n$  indivíduos em que  $k \leq n$  falhas distintas ocorrem nos tempos  $t_1, t_2, \dots, t_k$  e que a probabilidade condicional da  $i$ -ésima observação vir a falhar no tempo  $t_i$ , conhecendo as observações sob risco em  $t_i$ , é:

$$\frac{\alpha_i(t_i | x_i)}{\sum_{j \in R(t_i)} \alpha_j(t_i | x_j)} = \frac{\alpha_0(t_i) \exp\{\beta' x_i\}}{\sum_{j \in R(t_i)} \alpha_0(t_i) \exp\{\beta' x_j\}} = \frac{\exp\{\beta' x_i\}}{\sum_{j \in R(t_i)} \exp\{\beta' x_j\}} \quad (6)$$

$$L(\beta) = \prod_{i=1}^n \left( \frac{\exp\{\beta' x_i\}}{\sum_{j \in R(t_i)} \exp\{\beta' x_j\}} \right)^{\delta_i} \quad (7)$$

Em que os indivíduos censurados entram na função de verossimilhança parcial (7) por meio do conjunto de risco  $R_{(t_i)} = R_i$  e contribuem para esta função somente enquanto permanecem em risco. A estimação dos parâmetros  $\beta$  é feita maximizando-se o logaritmo da função de verossimilhança parcial (7) e resolvendo o sistema de equação obtido por essa maximização:

$$\frac{\partial l(\beta)}{\partial \beta_k} = \sum_{i=1}^n \delta_i \left[ x_{ik} - \frac{\sum_{j \in R_i} x_{jk} \exp\{\beta' x_j\}}{\sum_{j \in R_i} \exp\{\beta' x_j\}} \right] \quad k=1, \dots, p$$

em que  $l(\beta) = \log L(\beta)$ .

As propriedades assintóticas dos estimadores de máxima verossimilhança parcial -- necessárias para a construção de intervalos de confiança e testes de hipóteses --, foram estudadas por Andersen & Gill (1982) citados por Colosimo & Giolo (2006) e estas são consistentes e assintoticamente normais sob certas condições de regularidade. Pode-se, desse modo, fazer uso das estatísticas de Wald e da razão de verossimilhança para realizar inferências a partir do modelo de Cox.

## 2.8 Estimação de funções relacionadas com $\alpha_o(t)$

Não só coeficientes de regressão  $\beta$  são quantidades de grande interesse no modelo de Cox, como também as funções taxa de falha base acumulada  $H_o(t)$  e  $S_o(t)$  são também de grande importância por serem usadas em técnicas gráficas para avaliar a adequação do modelo ajustado. No entanto, essas quantidades não podem ser estimadas pelo método da máxima verossimilhança, pois o componente  $\alpha_o(t)$  não é especificado parametricamente. Como a função de verossimilhança parcial elimina o componente  $\alpha_o(t)$  completamente; logo, estimadores das quantidades mencionadas em (3) e (5) são de natureza não-paramétrica.

Segundo Colosimo & Giolo (2006), um estimador simples para a função de risco base acumulada  $H_o(t)$ , referenciado como estimador de Nelson-Aalen-Breslow ou estimador de Breslow, é uma função escada, com saltos nos distintos tempos de falha e expresso por:

$$\hat{H}_o(t_i) = \sum_{i: t_i \leq t} \frac{d(t_i)}{\sum_{j \in R_j} \exp(\hat{\beta}' x_j)}$$

em que  $d(t_i)$  é o número de falhas em  $t_i$ . Na ausência de covariáveis a expressão se reduz a

$$\hat{H}_o(t_i) = \sum_{i: t_i \leq t} \frac{d(t_i)}{n(t_i)}.$$

Uma vez estimada a função de risco base acumulada conforme Lawless (1982) obtêm-se a função de sobrevivência base estimada e a função de sobrevivência estimada para um indivíduo com vetor de covariáveis por meio das seguintes expressões, respectivamente:

$$\hat{S}_o(t) = \exp\{-\hat{H}_o(t)\} \text{ e } \hat{S}(t | x_i) = \left(\hat{S}_o(t)\right)^{\exp(\beta'x_i)}$$

### 2.8.1 Interpretação dos parâmetros no modelo de Cox

Os efeitos das covariáveis sobre a taxa de falha são medidos pelos coeficientes  $\beta$  no modelo de regressão de Cox, sendo que qualquer covariável pode acelerar ou desacelerar a função de risco.

A propriedade de riscos proporcionais é usada para interpretar os coeficientes estimados. Observe que a taxa de falha para dois indivíduos ( $i$  e  $l$ ) que apresentam os mesmos valores para as covariáveis, exceto para a  $p$ -ésima delas, é dada por:

$$\frac{\alpha(t | x_i)}{\alpha(t | x_l)} = \frac{\exp(\beta_p x_{ip})}{\exp(\beta_p x_{lp})} = \exp\{\beta_p (x_{ip} - x_{lp})\}$$

Por exemplo, para uma covariável dicotômica, como sexo, considere dois indivíduos, um do sexo masculino ( $x_{ip}$ ) e o outro do sexo feminino ( $x_{jp}$ ) e que a estimativa do parâmetro tenha sido 0,54 (valor meramente ilustrativo),

então a razão de risco entre os dois indivíduos será  $\frac{\alpha(x_{ip})}{\alpha(x_{jp})} = e^{0,54} = 1,72$ . Isto

quer dizer que o risco do sexo feminino para um evento é 1,72 vezes maior do que do sexo masculino, mantidas constantes as outras covariáveis que existiram.



## 2.9 Suposição de riscos proporcionais no modelo de Cox

O modelo de Cox é adequado para situações em que a suposição de riscos proporcionais é válida, ou seja, situações em que as funções de riscos não se cruzam.

Para verificar a suposição de riscos proporcionais do modelo de Cox, Therneau (2000) e Colosimo & Giolo (2006) consideram várias técnicas, entre elas, o método gráfico descritivo que se destaca pela sua simplicidade e visualização. Essa técnica consiste basicamente em: i) em dividir os dados em  $j$  estratos, usualmente de acordo com alguma covariável. ii) estimar  $\hat{H}_{oj}(t_i)$  para cada estrato  $j$  obtendo-se as curvas de  $\log \hat{H}_{oj}(t)$  versus  $t$ , ou  $\log(t)$ .

Se a suposição for válida, as curvas do logaritmo de  $\hat{H}_{oj}(t_i)$  versus  $t$  - ou  $\log(t)$  - devem apresentar diferenças aproximadamente constantes no tempo. Curvas não paralelas significam desvios da suposição de riscos proporcionais. É aconselhável construir este gráfico para cada covariável incluída no estudo. Se a covariável for de natureza contínua, uma sugestão é agrupá-la em um pequeno número de categorias. De acordo com Colosimo & Giolo (2006), uma vantagem desta técnica gráfica é a de indicar a covariável que está gerando a violação da suposição.

## 2.10 Generalização do modelo de regressão de Cox

Segundo Colosimo (2006), em algumas situações práticas envolvendo medidas longitudinais, o modelo de Cox não se ajusta adequadamente, pois podem existir covariáveis que, acompanhadas e monitoradas durante o tempo de estudo, podem mudar seus valores durante esse período. Uma das alternativas para enfrentar estas situações é uma extensão do modelo de Cox. Segundo Cox & Oakes (1984) e Collett (1994), o modelo de Cox pode ser estendido para

incorporar as informações longitudinais registradas para esta covariável. Outro problema pode estar no fato de o modelo não satisfazer a condição da suposição de riscos proporcionais. Para este caso, Lawless (1982) sugere um modelo alternativo, que é uma extensão do modelo de Cox, chamado de modelo de riscos proporcionais estratificado.

A análise estratificada consiste em dividir os dados de sobrevivência em estratos, de acordo com uma indicação de violação da suposição. O modelo de riscos proporcionais é então expresso como:

$$h_{ij}(t) = h_{oj}(t) \exp(x'_{ij}\beta) \quad (8)$$

para  $j = 1, \dots, m$  e  $i = 1, \dots, n_j$ , em que  $n_j$  é o número de observações no  $j$ -ésimo estrato.

Neste modelo (8), assume-se que os coeficientes de regressão são os mesmos em todos os estratos, embora o risco de base varie. A verossimilhança é estimada para cada indivíduo da mesma forma que na equação (7).

Colosimo & Giolo (2006) alertam que modelo estratificado só deve ser usado caso realmente seja necessário, como na presença de violação da suposição de riscos proporcionais, pois o uso desnecessário da estratificação pode acarretar perda de eficiência das estimativas obtidas.

### **2.11 Modelo de Cox estratificado**

O modelo de Cox estratificado é uma modificação do modelo de risco proporcional de Cox que leva em conta o controle pela estratificação dos preditores ou covariáveis que não satisfazem a suposição de risco proporcional. Preditores que satisfazem a suposição de risco proporcional são incluídos no modelo, enquanto os que não satisfazem, não são. Para esses preditores, que não

satisfazem a suposição de risco, não é possível obter os valores da razão de risco dos seus efeitos ajustados para os demais preditores, pois ela muda com o tempo: esse é o preço a ser pago pela estratificação.

As funções de riscos para os estratos serão diferentes somente se elas apresentarem funções de base diferentes, isto é,  $h_{o1}(t) \neq h_{o2}(t) \neq \dots \neq h_{on}(t)$ . No entanto, os coeficientes  $\beta_1, \beta_2, \dots, \beta_n$  serão iguais em todos os estratos. Como os  $\beta$ 's de cada preditor são os mesmos para diferentes estratos do(s) preditor(es) em que ocorre a violação da suposição de risco, então, estimativas como  $e^\beta$  são as mesmas para os estratos, essa característica é denominada de suposição da não interação do modelo de Cox estratificado e será detalhada mais tarde.

Formula-se uma função de verossimilhança para o modelo e para os dados a fim de que sejam obtidas as estimativas dos coeficientes dos parâmetros  $\beta$ 's. Essa função de verossimilhança para o modelo estratificado é diferente do modelo não estratificado. Para o modelo estratificado, a função de verossimilhança é obtida pela multiplicação das funções de verossimilhança de cada estrato. Assim, L (função de verossimilhança) é o produto de  $L_1.L_2...L_n$  em que sendo  $L_i$  com  $i=1,2,\dots,n$  são as funções de verossimilhança de cada um dos estratos (Kleinbaum & Klein, 1997).

De maneira geral, tem-se K covariáveis que violam a suposição de risco denominadas de  $Z_1, Z_2, \dots, Z_k$  e p covariáveis que não violam denominadas de  $X_1, X_2, \dots, X_p$ . Para então fazer análise da estratificação, cria-se uma nova covariável denominada  $Z^*$ , contendo as combinações das categorias de cada uma das covariáveis estratificadas. Essas combinações são os estratos da nova covariável  $Z^*$ . A fim de entender melhor o processo, considere a seguinte situação: tem-se duas covariáveis  $Z_1$  e  $Z_2$  que violam a suposição de riscos proporcionais, sendo  $Z_1$  uma variável contínua (idade, por exemplo) contendo três categorias A, B e C

e  $Z_2$  uma covariável binária. Desta forma, as seis combinações formadas com as covariáveis  $Z_1$  e  $Z_2$  compõem os estratos da covariável  $Z^*$ . Portanto a covariável  $Z^*$  terá  $K^*$  estratos, que são exatamente as combinações feitas com as categorias das covariáveis que violam a suposição de risco.

Já a forma geral da função de risco para o modelo estratificado é dada por  $h_g(t, X) = h_{og}(t) \exp[\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p]$ . Essa fórmula contém um índice  $g$  que indica os números dos estratos. Estratos esses que são definidos para as diferentes categorias da estratificação da covariável  $Z^*$  e é igual a  $K^*$  (número de combinações). Nota-se que as covariáveis  $X$ 's, que assumem a suposição de riscos proporcionais, são incluídas no modelo e a função de base  $h_{og}(t)$  é diferente para cada estrato e os coeficientes  $\beta$ 's são os mesmos para cada estrato.

## 2.12 O modelo sem interação

O modelo de Cox estratificado sem interação ocorre quando os  $\beta$ 's são os mesmos para os diferentes estratos. Para tornar o seu entendimento mais claro, considere o seguinte exemplo citado por Kleinbaum & Klein (1997, p.177). Toma-se uma amostra de tamanho  $n$  com três covariáveis A, B e C, em que a covariável A viola a suposição de risco proporcional e apresenta duas categorias  $A_1$  e  $A_2$ . A Tabela 4 mostra uma saída hipotética de um programa de computador da análise do modelo de Cox estratificado em que as covariáveis B e C não interagem com a covariável

TABELA 4 Estimativas dos parâmetros do modelo de Cox estratificado para covariável A sem interação com as covariáveis B e C, seus respectivos erros padrões (EP) e valor P

	<i>Coef</i> ( $\beta$ )	<i>EP</i>	$P >  z $
B	1,390	0,338	0,000
C	0,931	0,472	0,048

Log likelihood = -57.560

Os coeficientes  $\beta_1=1,390$  e  $\beta_2=0,931$  são iguais para os dois estratos de A. Se se considerar a interação das variáveis B e C com a variável A e ajustar-se separadamente o modelo para os estratos  $A_1$  e  $A_2$ , obtêm-se os valores apresentados nas Tabelas 5 e 6.

TABELA 5 Estimativas dos coeficientes, erros padrões e p-valor para modelo de Cox estratificado para covariável A (estrato A1) sem interação com as covariáveis B e C

	<i>Coef</i> ( $\beta$ )	<i>EP</i>	$P >  z $
B	1,639	0,519	0,002
C	1,859	0,729	0,011

Log verossimilhança = -22.100

TABELA 6 Estimativas dos coeficientes, erro padrão (EP) e p-valor para modelo de Cox estratificado para covariável A (estrato A2) sem interação com as covariáveis B e C

	<i>Coef</i> ( $\beta$ )	<i>EP</i>	$P >  z $
B	1,170	0,499	0,019
C	0,267	0,566	0,637

Log likelihood = -33.736

Tem-se agora o ajuste para os modelos com interação e sem interação. Quais dos dois modelos serão mais apropriados estatisticamente? Para responder

a essa pergunta deve-se fazer a comparação entre os dois modelos usando o teste da razão de verossimilhança e verificar se são ou não estatisticamente diferentes. Mas, antes disso, é necessário fazer um estudo detalhado da função de risco para os modelos com interação e sem interação. O modelo com interação será apresentado de duas formas: uma forma mais condensada denominada forma 1 (F1) e uma outra mais expandida, denominada forma 2 (F2).

No modelo sem interação, a função de risco assume a forma dada por  $h_g(t) = h_{og}(t) \exp[\beta_1 B + \beta_2 C]$  em que  $\beta_1$  assume valores iguais para os estratos  $A_1$  e  $A_2$ , assim como  $\beta_2$ . No entanto, a função risco de base assume valores diferentes para os estratos, em que  $g=1$  é para indicar o estrato  $A_1$  e  $g=2$  é para indicar o estrato  $A_2$ . Para o modelo com interação, tem-se a seguinte expressão para a função de risco para a F1  $h_g(t) = h_{og}(t) \exp[\beta_{1g} B + \beta_{2g} C]$ . Agora, cada preditor assume diferentes valores para os coeficientes nos estratos, assim como a função risco basal é também diferente. O índice  $g$  indica o coeficiente de cada estrato (Klein & Kleinbaum, 1997).

A forma F2, além de apresentar os efeitos principais, apresenta também os produtos das interações entre as covariáveis que violam e as que não violam a suposição de riscos proporcionais. Assim, a função de risco assume a seguinte forma para o exemplo citado:

$$h_g(t) = h_{og}(t) \exp[\beta_1 B + \beta_2 C + \beta_3 A.B + \beta_4 A.C]$$

em que

$$A = \begin{cases} 1 & \text{se } A_1 \\ 0 & \text{se } A_2 \end{cases}$$

Comparando as duas formas do modelo com interação, tem-se para  $g=1$ , o que implica  $A=1$ , a função de risco para F2 dada por

$$\begin{aligned} h_g(t) &= h_{o1}(t) \exp[\beta_1 B + \beta_2 C + \beta_3 A.B + \beta_4 A.C] \\ &= h_{o1}(t) \exp[(\beta_1 + \beta_3)B + (\beta_2 + \beta_4)C] \end{aligned} \quad (9)$$

A forma da função de risco para F1 é dada por

$$h_g(t) = h_{o1}(t) \exp[\beta_{11} B + \beta_{21} C] \quad (10)$$

Comparando-se (9) com (10), nota-se que

$$\beta_{11} = \beta_1 + \beta_3 \quad \text{e} \quad \beta_{21} = \beta_2 + \beta_4$$

Para  $g = 2$ , o que implica  $A=0$ , tem-se que a função de risco é dada para a F1 por

$$\begin{aligned} h_g(t) &= h_{o2}(t) \exp[\beta_1 B + \beta_2 C + \beta_3 0.B + \beta_4 0.C] \\ &= h_{o2}(t) \exp[\beta_1 B + \beta_2 C] \end{aligned} \quad (11)$$

A forma F1 para a mesma situação é dada por

$$h_g(t) = h_{o2}(t) \exp[\beta_{12} B + \beta_{22} C] \quad (12)$$

Comparando o modelo (11) com o modelo (12), tem-se que

$$\beta_1 = \beta_{12} \quad \text{e} \quad \beta_2 = \beta_{22}$$

Para ilustrar o exposto acima, considere novamente o exemplo hipotético apresentado anteriormente em que a saída hipotética de um programa de computador para o modelo estratificado com interação é dada pela Tabela 7.

TABELA 7 Estimativas dos coeficientes, erros padrões e p-valor para modelo de Cox estratificado para covariável A (estrato A1) com interação com as covariáveis B e C

	<i>Coef</i> ( $\beta$ )	<i>EP</i>	<i>P</i> >  <i>z</i>
B	1,170	0,499	0,019
C	0,267	0,566	0,637
AxB	0,469	0,720	0,515
AxC	1,592	0,923	0,084

Log likelihood=55.835

Comparando a Tabela 4 com a Tabela 6, nota-se que, para o estrato A<sub>1</sub>, o coeficiente da covariável B assume os valores para F1 de  $\beta= 1,639$  e, para a forma F2, o valor dado por  $\beta=1.170+0,469=1,639$ . Já o coeficiente para a covariável C, para o mesmo estrato, assume os valores para F1 de  $\beta= 1,859$  e para F2 de  $\beta=0,267+1,592=1,859$ . Nota-se que as duas formas na verdade são equivalentes, pois os coeficientes das covariáveis B e C para ambos os estratos são iguais.

Para responder a pergunta feita anteriormente, fez-se o teste da razão de verossimilhança. O teste comparou uma estatística para o modelo com interação e sem interação. A estatística usada foi  $LR=-2\ln L_R-(-2\ln L_F)$ , em que R é o modelo reduzido, neste caso, o modelo sem interação e F é o modelo completo, o modelo com interação. A estatística LR tem aproximadamente a distribuição de Qui-quadrado com dois graus de liberdade com a hipótese nula de que o modelo reduzido seja correto.



Voltando ao exemplo apresentado, tem-se que  $LR=115,120-111,670 = 3,45$ . Esse valor não é significativo ao nível de 5% para dois graus de liberdade, isto mostra que não existe diferença significativa entre os valores dos  $\beta$ 's para os dois estratos, logo não se rejeita a hipótese de que o modelo reduzido é apropriado, isto é, o modelo sem interação é aceitável.

Generalizando o que foi exposto, segundo Kleinbaum e Klein (1997), tem-se que:

Para o modelo em que há interação entre as covariáveis que não violam e as que violam a suposição de riscos proporcionais, a função de risco é dada por  $h_g(t, X) = h_{og}(t) \exp[\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p]$ . Nesse modelo, há várias covariáveis a serem estratificadas, no entanto, todas elas são reunidas em uma única covariável  $Z^*$ , em que seus estratos são combinações dos fatores de cada uma das covariáveis que as compõem. O índice  $g=1,2,3,\dots,K^*$  indica os estratos de  $Z^*$ .

Para o modelo em que há interação da covariável  $Z^*$  com as covariáveis  $X$ 's, têm-se duas formas, a saber:

A forma 1 (F1), em que os coeficientes de cada um dos fatores principais são diferente para cada extrato de  $Z^*$ . A função de risco é dada por  $h_g(t, X) = h_{og}(t) \exp[\beta_{1g} X_1 + \beta_{2g} X_2 + \dots + \beta_{pg} X_p]$ , em  $g = 1,2,3,\dots,K^*$  indica os diferentes extratos de  $Z^*$ .

A forma 2 (F2), que é mais extensa do que a forma 1, pois, além dos efeitos principais, contém os termos produtos envolvendo a covariável  $Z^*$  com cada covariável  $X$ 's. Nesse caso, a forma da função de risco é

$$\begin{aligned}
h_g(t, X) = h_{og}(t) \exp[ & \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \\
& + \beta_{11}(z_1^* X_1) + \beta_{21}(z_1^* X_2) + \dots + \beta_{p1}(z_1^* X_p) + \\
& + \beta_{12}(z_2^* X_1) + \beta_{22}(z_2^* X_2) + \dots + \beta_{p2}(z_2^* X_p) + \\
& + \dots + \beta_{1,k^*-1}(z_{k^*-1}^* X_1) + \dots + \beta_{p,k^*-1}(z_{k^*-1}^* X_p)]
\end{aligned}$$

Nota-se que a primeira linha contém os fatores principais e as linhas seguintes contém o produto de cada  $X_i$  com cada um dos estratos de  $Z^*$  respectivamente.

O teste estatístico realizado é o da razão de verossimilhança, que compara o modelo reduzido R (sem interação) com modelo completo F (com interação), tendo a estatística do teste uma distribuição de Qui-quadrado com  $p(k^*-1)$  graus de liberdade, com a hipótese nula de que cada termo produto do modelo com interação é igual a zero, isto equivale à hipótese de que o modelo sem interação é adequado. Assim, tem-se:

$$LR = -2 \ln L_R - (-2 \ln L_F)$$

em que R é modelo sem interação e F é o modelo com interação e mais

$LR : \chi^2_{p(k^*-1)}$  com a hipótese nula de

$$H_o = \begin{cases} \beta_{11} = \dots = \beta_{p1} = 0 \\ \beta_{12} = \dots = \beta_{p2} = 0 \\ \dots \\ \beta_{1,k^*-1} = \dots = \beta_{p,k^*-1} = 0 \end{cases}$$

### 2.13 Modelo de Cox estendido

Como visto anteriormente, a fórmula do modelo de Cox mostra que o risco num tempo  $t$  é o produto de duas quantidades. A primeira delas,  $h_o(t)$  é chamada de função de risco basal. A segunda quantidade é a expressão exponencial  $\exp(\sum \beta_i X_i)$ , em que  $\beta$ 's são os coeficientes de regressão e os  $X$ 's são as variáveis preditoras. Uma importante característica dessa fórmula é que  $h_o(t)$  não envolve os  $X$ 's, mas envolve  $t$ . Já os  $X$ 's, por não envolverem o tempo, são chamados de covariáveis tempo-independentes; se os  $X$ 's envolvessem o tempo, seriam denominados covariáveis tempo-dependente que são o objetivo de estudo dessa seção.

Se variáveis tempo-dependentes forem consideradas na análise, a forma do modelo de Cox poderá ser usada, mas o modelo não mais satisfará a suposição de riscos proporcionais e será chamado de modelo de Cox Estendido (Kleinbaum & Klein, 1997).

Para o modelo de riscos proporcionais de Cox, visto anteriormente, a forma geral para obter a razão de riscos proporcionais é dada por  $HR = \exp[\sum_i^n \hat{\beta}_i X_i^* - X_i]$  que compara duas quantidades de  $X$ 's, definidas como  $X^* = (X_1^*, X_2^*, \dots, X_p^*)$  e  $X = (X_1, X_2, \dots, X_p)$ . A suposição de risco nesse modelo é a de que o risco para um indivíduo é proporcional ao risco para um outro indivíduo, em que a constante de proporcionalidade mantém-se constante ao longo do tempo. Assim  $\frac{\hat{h}(t, X^*)}{\hat{h}(t, X)} = \hat{\Theta} \Rightarrow \hat{h}(t, X^*) = \hat{\Theta} \hat{h}(t, X)$ , em que  $\hat{\Theta}$  é a constante de proporcionalidade.

Agora, quando a variável tempo-dependente é usada para estimar a razão de riscos proporcionais para uma covariável tempo-independente, o

modelo de Cox é estendido para conter termos produtos envolvendo a covariável tempo-independente e uma função do tempo. Por exemplo, se a suposição de risco está sendo analisada para uma determinada variável A, o modelo de Cox pode ser estendido para incluir uma variável A.t adicionado a A. Se o coeficiente do termo produto não for significativo, conclui-se que a suposição de risco proporcional para variável A é satisfeita e que a escolha da variável A.t é uma escolha apropriada da covariável tempo-dependente.

Uma covariável tempo-dependente é definida como qualquer variável cujos valores podem mudar ao longo do tempo. Por exemplo, Raça é uma covariável tempo-independente, mas Religião é uma covariável tempo-dependente. Pode-se definir também uma covariável tempo-dependente como o produto de uma covariável tempo-independente vezes o tempo ou uma função do tempo. Por exemplo, a variável Raça, que é tempo-independente, a covariável definida pelo produto dela com o tempo (Raça x tempo) é tempo-dependente. Nesse exemplo, o tempo aparece sozinho, mas poderia ser dado por Raça x  $\log(t-4)$  ou Raça x  $g(t)$ . De uma maneira geral, tem-se  $E_{xg}(t)$ , em que E é uma variável qualquer, podendo assumir valores 0 ou 1 para definir o status em estudo e  $g(t)$  pode assumir valores iguais a 1, se o valor do tempo t for maior ou igual a um valor pré-definido  $t_0$  ou assumir 0 caso contrário, representando-se por

$$g(t) = \begin{cases} 1 & \text{se } t \geq t_0 \\ 0 & \text{se } t < t_0 \end{cases}.$$

Uma covariável tempo-dependente é considerada interna se a mudança de seu valor ao longo do tempo depender exclusivamente das características ou do comportamento específico do indivíduo, como exemplo, pode-se citar o nível de obesidade de uma pessoa. Caso contrário, se a mudança da variável depender

de influências externas, ela será chamada de “*ancillary*”. Como exemplo, pode-se citar o nível de poluição em uma determinada região que pode afetar vários indivíduos; o nível de desemprego que afeta vários indivíduos, pois são fatores que dependem somente de circunstâncias extrínsecas. Também, ela pode ser considerada em parte interna e em parte “*ancillary*”. Como exemplo, pode-se citar a covariável transplante cardíaco [HT(t)] associado a pessoas com sérios problemas cardíacos e candidatas a um transplante do coração. Essa variável assume valor 1 se a pessoa recebeu o transplante num tempo  $t_0$  inferior a  $t$  e um valor 0 se ela não recebeu o transplante e será considerada interna se as características individuais do paciente forem determinadamente importantes para decidir a realização ou não do transplante. Contudo, se houver incompatibilidade entre o doador e o paciente, a variável HT(t) é considerada “*ancillary*” (Kleinbaum & Klein, 1997).

Uma das dificuldades para decidir-se se uma covariável tempo-dependente poder ser classificada como “interna” ou “*ancillary*” é o fato de que rotinas de computador utilizadas para fazer análises estatísticas usando o modelo de Cox Estendido -- dependendo do programa de computador usado -- são diferentes para os dois tipos de variáveis. No entanto, independentemente da variável ser “interna” ou “*ancillary*”, a forma do modelo de Cox Estendido é a mesma e os procedimentos usados para obter estimativas dos coeficientes de regressão e os métodos para fazer inferência estatística são os mesmos que o modelo de riscos proporcionais de Cox.

Dada uma situação envolvendo ambas as variáveis preditoras tempo-independente e tempo-dependente, a forma da função de risco para incorporar ambos os tipos é dada por

$$h(t, X(t)) = h_o(t) \exp \left[ \sum_{i=1}^{p_1} \beta_i X_i + \sum_{j=1}^{p_2} \delta_j X_j(t) \right]$$



Já a razão de risco para o modelo de Cox Estendido é dada por

$$HR(t) = \frac{\hat{h}(t, X^*(t))}{\hat{h}(t, X(t))} = \exp \left[ \sum_{i=1}^{p_1} \hat{\beta}_i [X_i^* - X_i] + \sum_{i=1}^{p_2} \hat{\delta}_i [X_j^*(t) - X_j(t)] \right]$$

Sendo

$$X^*(t) = (X_1^*, X_2^*, \dots, X_{p_1}^*, X_1^*(t), X_2^*(t), \dots, X_{p_2}^*(t))$$

e

$$X(t) = (X_1, X_2, \dots, X_{p_1}, X_1(t), X_2(t), \dots, X_{p_2}(t))$$

Os componentes individuais para cada variável.

Uma importante característica do modelo de Cox Estendido é que este pode ser usado para verificar a suposição de risco proporcional de uma determinada covariável tempo-independente. Para isso, basta adicionar ao modelo de riscos proporcionais uma parcela contendo o produto da covariável tempo-independente com uma função do tempo, isto é, considere a  $i$ -ésima variável tempo-independente denotada por  $X_i$ , então define-se o  $i$ -ésimo termo produto como  $X_i \times g_i(t)$  em que  $g_i(t)$  é uma função de  $t$  para a  $i$ -ésima variável. Assim, a forma da função de risco para qualquer co-variável ou para um conjunto de covariáveis tempo-independente é dada por

$$h(t, X(t)) = h_o(t) \exp \left[ \sum_{i=1}^p \beta_i X_i + \sum_{i=1}^p \delta_i X_i g_i(t) \right]$$

Essa forma contém todas as covariáveis de interesse. O emprego desse modelo depende de uma decisão extremamente importante que é a forma da função  $g(t)$ , pois dependendo de sua escolha, a função de risco assume formas diferentes. Por exemplo:

Se  $g(t)=0$  para todas as covariáveis, tem-se o próprio modelo de risco proporcional, isto é,

$$h(t, X(t)) = h_o(t) \exp \left[ \sum_{i=1}^p \beta_i X_i \right]$$

Se  $g(t) = t \Rightarrow X_{ig_i(t)} = X_i t$ , então, para cada efeito principal, há um termo produto e a forma da função de risco é dada pela expressão

$$h(t, X(t)) = h_o(t) \exp \left[ \sum_{i=1}^p \beta_i X_i + \sum_{i=1}^p \delta_i X_i t \right]$$

Para verificar a suposição de risco somente para uma covariável tempo-independente, tem-se

$$g(t) = \begin{cases} t & \text{para a variável em questão} \\ 0 & \text{para qualquer outra variável} \end{cases}$$

e a forma da função de risco será

$$h(t, X(t)) = h_o(t) \exp \left[ \sum_{i=1}^p \beta_i X_i + \delta(Xt) \right]$$



Para  $g(t)$  igual a qualquer função, como por exemplo  $\ln(t)$ , tem-se  $g_i(t) = \ln t \Rightarrow X_i g_i(t) = X_i \ln t$  logo, a forma da função de risco será

$$h(t, X(t)) = h_o(t) \exp \left[ \sum_{i=1}^p \beta_i X_i + \sum_{i=1}^p \delta_i (X_i \ln(t)) \right]$$

E por último pode-se escolher  $g_i(t)$  igual a 1 para valores de  $t$  maior ou igual a um valor pré-determinado  $t_o$  ou  $g_i(t)$  igual a 0 se  $t$  for menor do que  $t_o$ .

Assim, tem-se  $g_i(t) = \begin{cases} 1 & \text{se } t \geq t_o \\ 0 & \text{se } t < t_o \end{cases}$ . Nesse caso, a função  $g(t)$  é denominada de “função *heaviside*”.

A verificação da suposição de risco proporcional para uma dada variável ou para um conjunto de covariáveis tempo-independente é feita considerando a hipótese nula de que o(s) valor(es) do(s) coeficiente(s)  $\delta$  das quantidades  $X_i g_i(t)$  dos termos produtos do modelo são iguais a zero, de modo que a hipótese fica  $H_o = \delta_1 = \delta_2 = \dots = \delta_p = 0$ .

Como no modelo de riscos proporcionais, os coeficientes de regressão no modelo estendido também são estimados usando os procedimentos da função de verossimilhança. As estimativas de máxima verossimilhança são obtidas maximizando a função de verossimilhança parcial e o teste estatístico usado para verificar a hipótese inicial  $H_o$  é o teste da razão de verossimilhança dado por  $LR = -2 \ln L_{PH} - (-2 \ln L_{\text{modelo de Cox estendido}})$ . Contudo, a parte computacional é mais complicada do que no modelo de riscos proporcionais, pois a construção do banco de dados é mais complicada quando as covariáveis são tempo-dependentes (Kleinbaum & Klein, 1997).

## 2.14 Adequação do ajuste do modelo de Cox

Devido à presença de censura na análise de sobrevivência e também pelo fato de os resíduos nem sempre apresentarem uma distribuição normal, podendo torná-los altamente assimétricos, diversos tipos de resíduos têm sido propostos na literatura para examinar o comportamento do modelo, principalmente sua adequação, o que pode ser feito usando técnicas gráficas. De acordo com Kleinbaum & Klein (1997), os principais resíduos são os resíduos de Cox, de Schoenfeld & Martingale.

Para uma amostra de  $n$  indivíduos, os resíduos proposto por Cox- Snell (Cox & Snell,1968), para o modelo de Cox, são definidos por:

$$e_i = \hat{H}_o(t_i) \exp \left\{ \sum_{k=1}^p x_{ip} \hat{\beta}_k \right\}, \quad i=1, \dots, n \quad (13)$$

Os quais, segundo Cox (1972), podem ser usados para verificar a qualidade de ajuste do modelo de Cox. Os resíduos de Cox-Snell são úteis para examinar o ajuste global do modelo de Cox, porém, quando o gráfico de  $\hat{H}(e_i)$  versus  $e_i$ , em que  $\hat{H}(e_i)$  é a função de risco acumulado obtida do modelo ajustado não se apresentar linear, os resíduos de Cox-Snell não podem detectar o tipo de falha apresentada pelo modelo.

De acordo com Therneau & Grambsch (2000), os resíduos propostos por Schoenfeld são usados para verificar a suposição de riscos proporcionais.

Considerando os  $k$  ( $k < n$ ) tempos distintos de falha  $t_1, t_2, \dots, t_k$  um vetor de resíduos de Schoenfeld é obtido em cada tempo observado de falha. Assim, se o indivíduo observado falhar, o correspondente resíduo é definido por:

$$r_i = x_i - \frac{\sum_{j \in R(t_i)} x_j \exp(\hat{\beta} x_j)}{\sum_{j \in R(t_i)} \exp(\hat{\beta} x_j)}. \quad (14)$$

Condicional a uma única falha no conjunto de risco  $R(t_i)$ .

A expressão  $\frac{\sum_{j \in R(t_i)} x_j \exp(\hat{\beta} x_j)}{\sum_{j \in R(t_i)} \exp(\hat{\beta} x_j)}$  é um estimador do valor esperado da

covariável.

Segundo Therneau (2000), os resíduos de Martingale resultam de uma modificação dos resíduos de Cox-Snell e são vistos como uma estimativa do número de falhas em excesso observada nos dados, mas não previsto pelo modelo. Assim, quando os dados apresentam censuras à direita e não forem dependentes do tempo e todas as covariáveis são fixadas no início do estudo, os resíduos martingale são definidos por:

$$\hat{m}_i = \delta_i - \hat{e}_i \text{ ou } \hat{m}_i = \delta_i - \hat{H}_o(t_i) \exp\left\{\sum_{k=1}^p x_{ip} \hat{\beta}_k\right\} = \delta_i - e_i, \quad i=1, \dots, n. \quad (15)$$

Em que,  $\delta_i$  é o indicador de censura, podendo assumir o valor 0, se o tempo de sobrevivência do indivíduo  $i$  é censurado; valor 1, se ele não é censurado e  $\hat{e}_i$  os resíduos de Cox-Snell.

Esses resíduos podem ser usados para verificar a adequação do modelo, mas em alguma situação eles podem ser usados para verificar a forma funcional das covariáveis e, dependendo da forma funcional, alguma transformação de dados tais como  $\log(x_i)$ ,  $x_i^2$ , pode ser usado no modelo para uma dada covariável.

Segundo Colosimo & Giolo (2006), os resíduos Deviance são definidos por:  $\hat{d}_i = \text{sinal}(\hat{m}_i)[-2(\hat{m}_i + \delta_i \log(\delta_i - \hat{m}_i))]^{1/2}$  em que  $\hat{m}_i$  é o resíduo de Martingale. O sinal de  $\hat{m}_i$  indica que o resíduo Deviance terá o mesmo sinal do resíduo de Martingale, sendo (+1), se for positivo e (-1), se for negativo. Para Colosimo & Giolo (2006), esse resíduo é uma tentativa de se fazer com que os resíduos Martingale sejam mais simétricos em torno de zero e que a plotagem dos gráficos dos resíduos Martingale em relação ao tempo ou Deviance versus tempos forneçam uma forma de verificar a adequação do modelo ajustado e auxiliem na detecção de observações atípicas.

Apesar de essa seção abordar os resíduos no modelo de Cox, a mesma abordagem é válida para o modelo de regressão paramétrico.

### **2.15 Modelos de regressão paramétricos**

Modelos de regressão paramétricos, modelos de locação e escala e modelos TFA são muitas vezes termos usados para definir uma classe de modelos denominadas de modelos de tempo de falha acelerado (TFA) ou *accelerated failure time models* (AFT). Lawless (1982) refere-se aos AFT como modelos de locação e escala ou modelos *log-linear*, que até então eram paramétricos. Portanto, esses três termos são usados na literatura como sinônimos. No entanto, hoje encontram-se modelos AFT em que a distribuição dos erros e, por consequência a distribuição do tempo de falha, não são especificados. Dessa forma, os termos AFT e modelos de regressão paramétricos não podem ser usados como sinônimos.

Nesse trabalho, apesar da ressalva feita acima, os termos modelos de regressão paramétricos, modelos de locação e escala, modelos *log-linear* e modelos AFT serão usados como sinônimos.

De acordo com Valença (1994), a classe de modelos de locação e escala é caracterizada pelo fato de  $Y = \log T$  ter uma distribuição com parâmetros de locação  $\mu(x)$  dependendo das covariáveis e um parâmetro de escala  $\sigma$ , constante, passando o modelo a ser definido por:

$$Y = \mu(x) + \sigma W \quad (16)$$

em que  $\sigma > 0$  e  $W$  tem uma distribuição que não depende de  $x$ .

Uma característica marcante desse modelo é que as covariáveis atuam multiplicativamente sobre  $T$ , enquanto o modelo de riscos proporcionais (PH) tem efeito multiplicativo no risco, como já foi dito anteriormente.

Para entender a pressuposição do AFT, considere o tempo de vida dos cães. Diz-se que os cães envelhecem 7 anos mais rápido do que os humanos, isto é, um cão de 10 anos de vida equivale a um ser humano de 70 anos. É o mesmo que dizer que a probabilidade de um cão sobreviver mais 6 anos é igual a probabilidade de um homem sobreviver mais 42 anos.

Se representar a função de sobrevivência do cão por  $S_1(t)$  e a do homem por  $S_2(t)$ , a suposição AFT pode ser representada por  $S_2(t) = S_1(\gamma \cdot t)$  para  $t \geq 0$ , em que  $\gamma$  é chamada de fator de aceleração entre o tempo de vida dos cães e dos homens. O fator de aceleração pode ser parametrizado como  $e^\alpha$  em que  $\alpha$  é um parâmetro a ser estimado para os dados. Com essa parametrização, a suposição AFT pode ser expressa como  $S_2(t) = S_1(e^\alpha \cdot t)$  ou  $S_1(t) = S_2(e^{-\alpha} \cdot t)$  para  $t \geq 0$ .

O fator de aceleração é a medida chave para associação obtida nos modelos AFT. Ele é o investigador para calcular o efeito da variável preditora sobre o tempo de sobrevivência ao comparar duas funções de sobrevivência. Ele apresenta a mesma importância que a razão de risco nos modelos de riscos proporcionais.

De acordo com o conjunto de dados  $\boldsymbol{\mu}(\mathbf{x})$ , pode assumir várias formas funcionais. Entre elas, a forma  $\boldsymbol{\mu}(\mathbf{x})=\mathbf{x}\boldsymbol{\beta}$  em que  $\mathbf{x} = (x_1, \dots, x_{p-1})$  é o vetor  $1 \times (p-1)$  de covariáveis, que podem ser quantitativas ou qualitativas e  $\boldsymbol{\beta} = (\beta_0, \dots, \beta_{p-1})^T$  é um vetor  $p \times 1$  de parâmetros desconhecidos. Utilizando esta suposição, modelo de locação e escala é um modelo log-linear para  $T$  com resíduo  $W$ .

A função de sobrevivência de  $Y$  dado  $\mathbf{x}$  é da forma  $G\left(\frac{y - \mu(x)}{\sigma}\right)$ , em que  $G(\cdot)$  é a função de sobrevivência de  $W$ , sendo expressa da seguinte forma:

$$S(t | x) = G\left[\log(t | \alpha(x))^\delta\right] \text{ em que } \alpha(x) = e^{\mu(x)} \text{ e } \delta = \frac{1}{\sigma}.$$

A seguir, serão apresentados os modelos de regressão paramétricos usados no trabalho.

## 2.16 Modelo de regressão exponencial

Se  $T$  dado  $\mathbf{x}$  tem distribuição exponencial, seja  $Y=\log T$ . O modelo de regressão linear para  $T$  é dado por:

$$Y = \mathbf{x}\boldsymbol{\beta} + W$$

Em que  $W$  tem distribuição valor extremo padrão que é da seguinte forma:

$$f(w) = \exp[w - \exp(w)], \quad w \in \mathfrak{R}.$$

Sendo que, a f.d.p. e a função de sobrevivência de  $Y$  dado  $\mathbf{x}$  são, respectivamente:

$$f_y(y|x) = \exp[y - x\beta - \exp(y - x\beta)], y \in \mathfrak{R}$$

e

$$S_y(y|x) = \exp[-\exp(y - x\beta)]$$

### 2.17 Modelo de regressão Weibull

Se  $T$  dado  $\mathbf{x}$  tem distribuição Weibull com parâmetros  $\alpha(x) = e^{x\beta}$  e  $\beta = \frac{1}{\sigma}$ , o modelo de regressão log-linear é da forma:

$Y = x\beta + \sigma W$ . Em que  $W$  tem distribuição valor extremo padrão. A f.d.p e a função de sobrevivência para  $Y$  dado  $\mathbf{x}$  são, respectivamente:

$$f_y(y|x) = \frac{1}{\sigma} \exp\left[\left(\frac{y - x\beta}{\sigma}\right) - \exp\left(\frac{y - x\beta}{\sigma}\right)\right]_{y \in \mathfrak{R}},$$

$$S_y(y|x) = \exp\left[-\exp\left(\frac{y - x\beta}{\sigma}\right)\right]$$

### 2.18 Modelo de regressão Log-normal

O modelo de regressão paramétrico, cujo tempo de vida  $T$  dado  $\mathbf{x}$  tem distribuição Log-Normal, pode ser representado por:

$$Y = \mathbf{x}\beta + \sigma Z$$

Em que  $Z$  tem distribuição Normal Padrão ( $\phi(z)$ ). A f.d.p. e a função de sobrevivência para este modelo são, respectivamente:

$$f_y(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{y-x\beta}{\sigma}\right)^2\right]_{y \in \mathfrak{R}}$$

e

$$S_y(y|x) = 1 - \Phi\left(\frac{y-x\beta}{\sigma}\right),$$

Sendo a Normal Acumulada Padrão definida por

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$$

## 2.19 Modelo de regressão Gama generalizada

O modelo de regressão paramétrico, cujo tempo de vida  $T$  dado  $x$  tem distribuição Log-Gama generalizada, pode ser representado por:

$$Y = \mathbf{x}\boldsymbol{\beta} + \sigma Z$$

Em que,  $Z$  tem distribuição dada por:

$$f_z(z) = \begin{cases} \frac{|q|}{\Gamma(q^{-2})} (q^{-2}) \exp\{q^{-1}z - q^{-2} \exp(qz)\}, & q \neq 0 \\ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right), & q = 0 \end{cases}$$



A função de sobrevivência de Y dado x é representada por:

$$S_Y(y|x) = \begin{cases} Q(q^{-2}, q^{-2} \exp\{q(y-x\beta)/\sigma\}) & q > 0 \\ 1 - Q(q^{-2}, q^{-2} \exp\{q(y-x\beta)/\sigma\}) & q < 0 \\ 1 - \Phi[(y-x\beta)/\sigma] & q = 0 \end{cases}$$

Sendo  $Q(k, a) = \int_a^{\infty} \frac{x^{k-1} e^{-x}}{\Gamma(k)} dx$ , a função gama incompleta.

## 2.20 Inferência nos modelos de regressão paramétrico

Para avaliar e comparar modelos de regressão paramétricos de sobrevivência, pode-se utilizar o teste de Wald e a análise da função desvio, também chamada teste da razão de verossimilhança (esse teste foi utilizado nesse trabalho).

A estatística de Wald testa a hipótese de que o parâmetro  $\beta$  da regressão é igual a zero.

$$z = \frac{\hat{\beta}}{ep(\hat{\beta})}$$

Em que  $ep(\hat{\beta})$  é o erro padrão de  $\hat{\beta}$ .

A análise de Deviance pode ser utilizada tanto para obter uma estatística global do ajuste do modelo como para comparar modelos. A hipótese nula  $H_0$ , de que o modelo se ajusta aos dados, pode ser testada pela estatística:

$$D = 2(l_{\text{modelo}} - l_{\text{nulo}})$$

Em que  $l_{\text{modelo}}$  é o logaritmo da função de verossimilhança do modelo com as covariáveis e  $l_{\text{nulo}}$  é o logaritmo da função de verossimilhança do modelo nulo (modelo sem covariáveis). Essa estatística segue, assintoticamente, uma distribuição Qui-quadrado com número de graus de liberdade igual ao número de covariáveis do modelo.

### 3 MATERIAL E MÉTODOS

O propósito dessa seção é descrever os procedimentos metodológicos usados na pesquisa em que foram entrevistadas mulheres que tiveram seus filhos na Maternidade da Santa Casa de Misericórdia de Campo Belo, com auxílio de um questionário construído com a finalidade de avaliar os fatores que contribuem para a ocorrência da gravidez na adolescência.

Campo Belo, cidade do Estado de Minas Gerais, abrange uma área de 531 quilômetros quadrados. Situa-se à margem direita do Rio Grande e dista da capital do estado, Belo Horizonte, a 232 quilômetros por rodovia asfaltada.

#### 3.1 Material

Os dados para a realização da pesquisa foram obtidos na Maternidade da Santa Casa de Misericórdia de Campo Belo, Estado de Minas Gerais, que atende as seguintes localidades: Campo Belo, Cristais, Santana do Jacaré, Aguanil e Dias.

A população objeto de estudo foram todas as mulheres que deram à luz na Maternidade da Santa Casa de Misericórdia São Vicente de Paula, localizada na cidade de Campo Belo, no período compreendido entre os anos de 2000 e 2006.

Inicialmente, fez-se um levantamento junto ao banco de dados da Secretaria Municipal de Saúde de Campo Belo e constatou-se que de 2000 até 2006, 4800 gestantes deram à luz na Maternidade. Dessas, 1440 eram adolescentes.

A coleta dos dados das participantes foi realizada no período de janeiro de 2007 a setembro de 2008, coletando todos os registros armazenados no banco de dados do hospital referentes às gestantes que deram à luz no período de janeiro de 2000 a janeiro de 2006.

O processo de coleta dos dados pode ser resumido da seguinte maneira: inicialmente, foram xerografados os livros de registro da maternidade. De posse do nome da gestante, digitava-se o mesmo no computador e anotavam-se em uma ficha, os dados de interesse. De cada gestante, foram levados em consideração apenas os dados relativos à primeira gravidez notificada no período de 2000 a 2006, não sendo incluídos na pesquisa os dados relativos a qualquer outra gestação posterior.

Foram usados como critérios de exclusão: a não-aceitação do termo de consentimento, faixa etária inferior a 12 anos e a impossibilidade de contactação da participante.

A distribuição dos questionários ocorreu da seguinte maneira: em Campo Belo, o questionário foi entregue à enfermeira chefe de cada PSF, quem fez uma leitura deste, do termo de aceite, interagiu com o objetivo da pesquisa e discutiu os pontos duvidosos com o pesquisador. Em seguida, os envelopes contendo a ficha de endereço, o questionário e o termo de aceite para que pudesse ser feita a distribuição foram repassados aos agentes de saúde.

Nas cidades vizinhas, as mulheres que faziam parte da amostra foram convidadas a comparecer aos PSFs e lá receberam as fichas com seus nomes e o questionário para que pudessem respondê-los ali mesmo.

À medida que os questionários iam sendo recolhidos, estes eram organizados em uma planilha eletrônica. O nome das mulheres que participaram voluntariamente deste estudo foi omitido, somente o número foi processado e a ficha que contém o nome e o endereço da participante foi incinerada ou ficou em poder dela. Os dados foram digitados e conferidos duas vezes pelo próprio pesquisador na tentativa de evitar qualquer erro de digitação.

## **3.2 Método**

### **3.2.1 Definição do evento**

O evento de interesse será a data em que a mulher ficou grávida após os 12 anos de idade. O início do tempo de estudo será o de 12 anos e o final, 18 anos de idade cronológica, apesar de muitos psicólogos divergirem quanto ao início e fim da fase de adolescência. Nesse projeto, serão consideradas adolescentes todas as mulheres que estiverem na faixa de 12 aos 18 anos. Essa escolha se deu devido ao fato de que no levantamento feito junto a Secretaria Municipal de Saúde, constatou-se que o número de meninas grávidas na faixa abaixo de 12 anos era muito pequeno, cerca de 3 no período de 2000 a 2006, sendo considerados casos atípicos, portanto excepcionais para os fins deste estudo. No entanto, o número de meninas grávidas na faixa de 12 a 18 anos foi muito significativo, aproximadamente 1200 adolescentes no período de 2000 a 2006.

O evento de interesse, data em que ocorreu a gravidez, foi medido em semanas ou dias da seguinte maneira: toma-se a data em que a gestante deu à luz em semanas e subtrai-se 40,0, que correspondem aos meses de gestação.

Do resultado obtido em 1, subtrai-se 648, correspondentes a 12 anos, em semanas.

O valor obtido é aproximadamente o tempo do evento de interesse. Se o valor encontrado for menor do que 324, então ocorreu a falha, pois esse período de 324 semanas corresponde à fase considerada como adolescente. Caso contrário, terá ocorrida a censura, isto é, a mulher ficou grávida após os 18 anos de idade e, como exemplos, tem-se:

Considerando que a gestante deu à luz aos 17 anos e 3 meses; isto equivale a 931,5 semanas de vida. Subtraindo 40,0, tem-se 891,5 semanas, indicando que ela ficou grávida aos 16 anos e 6 meses. Subtraindo de 891,5 o

valor 648, tem-se 243,5 semanas, que é o tempo do evento de interesse, indicando que ela ficou grávida 243,5 semanas após os 12 anos de idade, que é considerado o início do tempo de estudo. Nesse caso, ocorreu a falha ou evento, pois o valor 243,5 é inferior a 324.

Considerando que a gestante deu à luz aos 20 anos e 6 meses; isto equivale a 1107,0 semanas de vida. Subtraindo 40,0, tem-se 1067,0 semanas. Logo, ela ficou grávida aos 19 anos e 9 meses. Subtraindo de 1067,0 o valor 648,0, tem-se 419,0 semanas, que é o tempo do evento de interesse, indicando que ela ficou grávida 419,0 semanas após os 12 anos de idade, início do tempo de estudo. Nesse caso, ocorreu a censura, pois o valor 419,0 é superior a 324.

### 3.2.2 Descrição das covariáveis.

As covariáveis a serem estudadas e como foram avaliadas suas categorias estão apresentadas na Tabela 8.

TABELA 8 Descrição das covariáveis e das categorias que compõem o banco de dados usado no estudo da gravidez na adolescência na região de Campo Belo, MG

<i>Covariáveis</i>	<i>Categorias</i>
1. Renda per capita familiar (em salário mínimo/mês)	1 ou mais/ menor que 1
2. Idade da adolescente	Abaixo de 18 / acima
3. Morou com apenas um dos pais durante a adolescência	Sim / não
4. Escolaridade do pai no instante em que ficou grávida	Até o 1º grau/ mais que o 1º grau
5. Escolaridade da mãe no instante em que ficou grávida	Até o 1º grau/ mais que o 1º grau
6. Escolaridade da adolescente	Até o 1º grau / mais que o 1º grau

(Continua...)

“TABELA 8, Cont.”

<i>7. Religião</i>	<i>Católica / evangélica</i>
8. Assiste à TV (filmes, novelas, etc)	Sim / Não
9. Tem vida social ativa na adolescência	Sim / Não
10. Diálogo com os pais sobre sexo, etc.	Sim / Não
11. Morou com ambos os pais na adolescência	Sim / Não
12. Raça, quanto à afro-descendência	Sim / Não
13. Fez uso de alguma droga (álcool, etc)	Sim / Não
14. Teve orientação sexual na escola	Sim / Não
15. Tem conhecimento dos métodos contraceptivos	Sim / Não
16. Facilidade em obter anticoncepcionais	Sim / Não
17. Teve vida sexual ativa	Sim / Não
18. Conhece as responsabilidades legais do parceiro com o filho	Sim / Não
19. Conhece os programas sociais de apoio à gravidez	Sim / Não

### 3.3 Análise dos dados

Como as covariáveis eram todas dicotômicas, fez-se inicialmente uma comparação entre suas categorias usando as estimativas de Kaplan- Meier com o objetivo de verificar se existiam diferenças entre suas curvas de sobrevivência. Como o número de covariáveis envolvidas foi grande e, segundo Collett (1994), fica inviável ajustar todos os possíveis modelos formados pela combinação de todas estas covariáveis para depois selecionar o melhor modelo, fez-se uma seleção das covariáveis que podem ser utilizadas para descrever o comportamento da variável resposta.

Para a seleção das covariáveis, utilizou-se inicialmente os testes de log-rank e Wilcoxon, como ponto de partida. Depois, utilizou-se da estratégia derivada da proposta de Collett (1994) e detalhada em Colosimo & Giolo (2006, p 146), os quais sugeriram os seguintes passos:

Ajustar todos os modelos contendo uma única variável e incluir no modelo todas as covariáveis que forem significativas no nível de 0,10. Será utilizado o teste da razão de verossimilhança neste passo;

Ajustar conjuntamente as covariáveis significativas selecionadas no passo 1. Na presença de certas covariáveis, outras podem deixar de ser significativas. Consequentemente, ajustam-se modelos reduzidos, excluindo uma única variável de cada vez. Verificam-se as covariáveis que provocam um aumento estatisticamente significativo na estatística da razão de verossimilhança. Somente aquelas que atingirem a significância permanecem no modelo;

Ajustar um novo modelo com covariáveis retidas no passo 3. Neste passo, as covariáveis excluídas no passo 2 retornam ao modelo para confirmar que elas não são estatisticamente significativas;

Incluir no modelo as possíveis covariáveis significativas no passo 3 juntamente com aquelas do passo 2. Neste passo, as covariáveis excluídas no passo 1 são retornadas para confirmar se elas são ou não significativas;

Ajustar um modelo incluindo as covariáveis significativas no passo 4. Neste passo, testa-se ainda se algumas delas podem ser retiradas do modelo;

Utilizar as covariáveis que permaneceram após o passo 5, ajustando o modelo final para os efeitos principais. Para completar a modelagem, deve-se verificar a possibilidade de inclusão de termos de interação. Testa-se cada uma das interações duas a duas entre as possíveis covariáveis já incluídas no modelo. O modelo final ficará determinado pelos efeitos principais identificados no passo 5 e os termos de interação significativos identificados neste passo.

A cada passo do processo de seleção de variáveis, a estatística do teste é obtida utilizando-se o teste da razão de verossimilhança com a estatística, seguindo uma distribuição Qui-quadrado de referência, com graus de liberdade igual ao número de termos excluídos (diferença entre o número de parâmetros dos dois modelos a serem comparados).



Após a seleção do grupo de covariáveis, foram utilizados os modelos de regressão de locação e escala e os modelo de Cox (riscos proporcionais, estratificados e variáveis tempo-dependentes) na análise dos dados.

Para se fazer o ajuste do modelo de locação e escala, com o objetivo de escolher o modelo ajustado para interpretar as estimativas das covariáveis, foram comparados os modelos Exponencial, Weibul e Log-Normal, cujas distribuições de probabilidade foram detalhadas seção 2.20. Partiu-se do pressuposto de que o modelo gama generalizado é adequado aos dados de sobrevivência em estudo. Após escolher o conjunto de covariáveis que compuseram o modelo, foi verificada a utilização de modelos mais simples, casos especiais da gama generalizada, mas não menos adequados aos dados. O teste da razão de verossimilhança foi utilizado para selecionar os modelos.

A adequação do modelo foi verificada por meio da análise de resíduos. Só assim foi feita a interpretação das estimativas das quantidades desconhecidas do modelo.

A estimação e teste dos parâmetros dos modelos foram feitos utilizando-se os estimadores de máxima verossimilhança.

Quanto ao modelo de Cox, definido após a escolha do conjunto de covariáveis que o compuseram, verificou-se a violação da suposição de riscos proporcionais. Como esta suposição não foi satisfeita para algumas das covariáveis, foram então ajustados os modelos de Cox estratificado e o Cox com covariável tempo-dependente, que são generalizações do modelo de Cox.

As análises dos dados foram realizadas utilizando o Sistema Computacional Estatístico R, basicamente a '*Library Survival*' de Terry Therneau, original R port by Thomas Lumley, conforme R Development Core Team (2008).

## 4 RESULTADOS E DISCUSSÃO

### 4.1 Análise descritiva e exploratória dos dados

Segundo Colosimo & Giolo (2006), a primeira etapa de qualquer análise estatística de dados consiste de uma análise descritiva das covariáveis envolvidas no estudo. No entanto, por se tratar de análise de sobrevivência, foram usados métodos não-paramétricos para cada covariável. Como as covariáveis são dicotômicas, como mostra a Tabela 8 na seção 3.5 suas categorias foram comparadas usando as estimativas de Kaplan & Méier (1958). Como exemplo, observe a Figura 1, que mostra as curvas de Kaplan & Méier (1958) para a covariável Religião. Nota-se que as mulheres de religiões evangélicas engravidaram mais tarde, isto é, apresentaram um tempo para engravidar maior do que as católicas. As mulheres da religião católica representam 79,06% do total das mulheres entrevistadas.

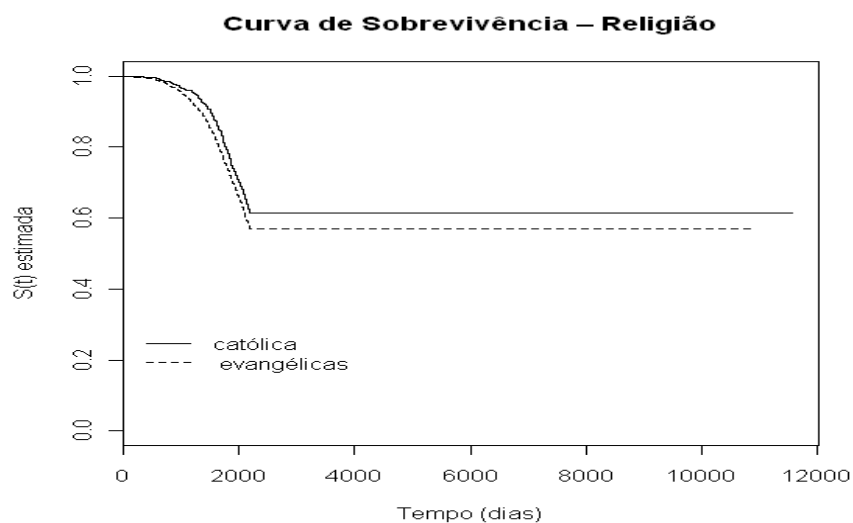


FIGURA 1 Representação gráfica das curvas de sobrevivência estimadas pelo método de Kaplan & Méier (1958) para a covariável religião católica e evangélicas.

## 4.2 Seleção do modelo paramétrico

Por se tratar de estudo envolvendo muitas covariáveis, algumas delas não fizeram parte do modelo, devido à sua contribuição nula ou pequena. Para selecionar as covariáveis que compuseram o modelo adequado, todas as covariáveis foram modeladas pelo teste de logrank e Wilcoxon, permanecendo aquelas com p-valor inferior a 0,25, de acordo com Colosimo & Giolo (2006). A Tabela 9 mostra os valores dos testes de logrank e de Wilcoxon para as 19 covariáveis em estudo. Nota-se que as covariáveis média (M), escolaridade do pai (EP), escolaridade da mãe (EM) e diálogo(D) com os pais sobre sexo estão sujeitas a deixar o processo de seleção por apresentarem valores p superiores a 0,25. Essas covariáveis -- que deixaram a seleção do modelo nesse estágio -- voltarão mais à frente, em outros passos, para verificar a sua relevância naquele estudo.

TABELA 9 Estimativas dos testes logrank e Wilcoxon e respectivas probabilidades (p-valor) utilizados para testar a igualdade das curvas de sobrevivência para as diversas covariáveis usadas no estudos da gravidez na adolescência na microrregião de Campo Belo - MG

Covariáveis	Testes	
	logrank	Willcoxon
Ren: Renda per capita familiar	9,1947 (0,0268)	9,6570 (0,0217)
Rel: Religião	6,2451 (0,0125)	7,9134 (0,0049)
M: Mídia (Assiste à TV )	0,0651 (0,7986)	0,1717 (0,6786)
Cor: Cor da pele	4,7853 (0,0914)	5,6987 (0,0579)
EP: Escolaridade do pai	4,8880 (0,2990)	3,7310 (0,4436)
EM: Escolaridade da mãe	3,2167 (0,5222)	3,3959 (0,4939)
E: Escolaridade da Adolescente	28,01(<,0001)	22,65 (0,0001)
MP: Morou com um(dois) dos pais	3,4486 (0,0633)	3,8654 (0,0493)
PS: Os pais se separaram na adolescência	30,1961 (<,0001)	28,9834 (<,0,001)
OS: Teve orientação sexual durante na escola	27,7636 (<,0001)	42,0199 (<,0,001)
VSA: Vida sexual ativa na adolescência	146,989 (<,0001)	146,671 (<,0,001)

(Continua...)

“TABELA 9, Cont.”

RL: Responsabilidade legal do parceiro	20,4333 (<,0001)	32,7275 (<0,001)
PA: programas de assistências	16,1240 (<,0001)	13,3720 (0,0003)
MC: métodos contraceptivos	379,1391(<,0001)	497,091 (<,0001)
VSOA: Vida social ativa	89,8878 (<,0001)	93,9566 (<,0001)
D: Diálogo com os pais sobre sexo, etc.	0,0000 (0,9945)	0,0026 (0,9592)
Dr: Fez uso de drogas na adolescência	10,1606 (0,0014)	9,7042 (0,0018)
AO: Facilidade em obter anticoncepcionais	2,1325 (0,1442)	1,1880 (0,2757)

A Tabela 10 mostra os modelos usados nos na seleção do modelo paramétrico.

TABELA 10 Número do modelo e discriminação do modelo usados para obter o modelo paramétrico ajustado aos dados do estudo da gravidez na adolescência na microrregião de Campo Belo- MG

<i>Número do Modelo</i>	<i>Modelo</i>
1	Nulo
2	Ren
3	Rel
4	Cor
5	MP
6	PS
7	OS
8	VSA
9	RL
10	PA
11	MC
12	VSOA
13	Dr
14	AO
15	Ren+Rel+Cor+MP+PS+OS+VSA+RL+PA+MC+VSOA+Dr

(Continua...)

“TABELA 10, Cont.”

---

16	<i>Rel+Cor +MP+PS+OS+VSA+RL+PA+MC+VSOA+Dr</i>
17	Ren+Cor +MP+PS+OS+VSA+RL+PA+MC+VSOA+Dr
18	Ren+Rel +MP+PS+OS+VSA+RL+PA+MC+VSOA+Dr
19	Ren+Rel+Cor +PS+OS+VSA+RL+PA+MC+VSOA+Dr
20	Ren+Rel+Cor +MP+OS+VSA+RL+PA+MC+VSOA+Dr
21	Ren+Rel+Cor +MP+PS+VSA+RL+PA+MC+VSOA+Dr
22	<i>Ren+Rel+Cor +MP+PS+OS+RL+PA+MC+VSOA+Dr</i>
23	Ren+Rel+Cor +MP+PS+OS+VSA+PA+MC+VSOA+Dr
24	Ren+Rel+Cor +MP+PS+OS+VSA+RL+MC+VSOA+Dr
25	Ren+Rel+Cor +MP+PS+OS+VSA+RL+PA+VSOA+Dr
26	Ren+Rel+Cor +MP+PS+OS+VSA+RL+PA+MC+Dr
27	Ren+Rel+Cor +MP+OS+OS+VSA+RL+PA+MC+VSOA
28	Rel +Cor+OS+VSA+RL+PA+VSOA+Dr
29	Cor+OS+VSA+RL+PA+VSOA+Dr
30	Rel +OS+VSA+RL+PA+VSOA+Dr
31	Rel +Cor+VSA+RL+PA+VSOA+Dr
32	Rel +Cor+OS+RL+PA+VSOA+Dr
33	Rel +Cor+OS+VSA+PA+VSOA+Dr
34	Rel +Cor+OS+VSA+RL+VSOA+Dr
35	Rel +Cor+OS+VSA+RL+PA+Dr
36	Rel +Cor+OS+VSA+RL+PA+VSOA
37	Rel +OS+VSA+RL+VSOA
38	OS+VSA+RL+VSOA
39	Rel +VSA+RL+VSOA
40	Rel +OS+RL+VSOA
41	Rel +OS+VSA+VSOA
42	Rel +OS+VSA+RL
43	Rel +OS+VSA+RL+VSOA
44	Rel +OS+VSA+RL+VSOA+M
45	Rel +OS+VSA+RL+VSOA+EP
46	Rel +OS+VSA+RL+VSOA+EM
47	Rel +OS+VSA+RL+VSOA+E
48	Rel +OS+VSA+RL+VSOA+D
49	Rel +OS+VSA+RL+VSOA+EP+E+D

(Continua...)

“TABELA 10, Cont.”

---

50	OS+VSA+RL+VSOA+EP+E+D
51	Rel +VSA+RL+VSOA+EP+E+D
52	Rel +OS+RL+VSOA+EP+E+D
53	Rel +OS+VSA+VSOA+EP+E+D
54	Rel +OS+VSA+RL+EP+E+D
55	Rel +OS+VSA+RL+VSOA+E+D
56	Rel +OS+VSA+RL+VSOA+EP+D
57	Rel +OS+VSA+RL+VSOA+EP+E
58	Rel +OS+VSA+RL+VSOA
59	Rel +OS+VSA+RL+VSOA+Ren
60	Rel +OS+VSA+RL+VSOA+Cor
61	Rel +OS+VSA+RL+VSOA+MP
62	Rel +OS+VSA+RL+VSOA+PS
63	Rel +OS+VSA+RL+VSOA+MC
64	Rel +OS+VSA+RL+VSOA+Ren+MP+MC+PS
65	OS+VSA+RL+VSOA+Ren+MP+MC+PS
66	Rel+VSA+RL+VSOA+Ren+MP+MC+PS
67	Rel +OS+RL+VSOA+Ren+MP+MC+PS
68	Rel +OS+VSA+VSOA+Ren+MP+MC+PS
69	Rel +OS+VSA+RL+Ren+MP+MC+PS
70	Rel +OS+VSA+RL+VSOA+MP+MC+PS
71	Rel +OS+VSA+RL+VSOA+Ren+MC+PS
72	Rel +OS+VSA+RL+VSOA+Ren+MP+PS
73	Rel +OS+VSA+RL+VSOA+Ren+MP+MC
74	Rel +OS+VSA+RL+VSOA+E+D+Ren+PS +MC
75	OS+VSA+RL+VSOA+E+D+Ren+PS +MC
76	Rel +VSA+RL+VSOA+E+D+Ren+PS +MC
77	Rel +OS+RL+VSOA+E+D+Ren+PS +MC
78	Rel +OS+VSA+VSOA+E+D+Ren+PS +MC
79	Rel +OS+VSA+RL+E+D+Ren+PS +MC
80	Rel +OS+VSA+RL+VSOA+D+Ren+PS +MC
81	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC
82	Rel +OS+VSA+RL+VSOA+E+D+PS +MC
83	Rel +OS+VSA+RL+VSOA+E+D+Ren +MC
84	Rel +OS+VSA+RL+VSOA+E+D+Ren+PS
85	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC
86	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+Rel*OS

(Continua...)

“TABELA 10, Cont.”

---

87	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+Rel*VSA
88	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+Rel*RL
89	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+Rel*VSOA
90	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+Rel*E
91	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+Rel*Ren
92	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+Rel*PS
93	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+Rel*MC
94	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+OS*VSA
95	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+OS*RL
96	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+OS*VSOA
97	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+OS*E
98	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+OS*Ren
99	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+OS*PS
100	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+OS*MC
101	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+VSA*RL
102	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+VSA*VSOA
103	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+VSA*E
104	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+VSA*Ren
105	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+VSA*PS
106	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+VSA*MC
107	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+RL*VSOA
108	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+RL*E
109	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+RL*Ren
110	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+RL*PS
111	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+RL*MC
112	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+VSOA*E
113	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+VSOA*Ren
114	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+VSOA*PS
115	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+VSOA*MC
104	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+VSA*Ren
105	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+VSA*PS
106	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+VSA*MC
107	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+RL*VSOA
108	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+RL*E
109	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+RL*Ren
110	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+RL*PS

(Continua...)

“TABELA 10, Cont.”

111	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+RL*MC
112	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+VSOA*E
113	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+VSOA*Ren
114	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+VSOA*PS
115	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+VSOA*MC
116	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+E*Ren
117	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+E*PS
118	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+E*MC
119	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+Ren*PS
120	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+Ren*MC
121	Rel +OS+VSA+RL+VSOA+E+Ren+PS +MC+PS*MC
122	Rel +OS+VSA+RL+VSOA+E+Ren+PS
123	OS+VSA+RL+VSOA+E+Ren+PS
124	Rel +VSA+RL+VSOA+E+Ren+PS
125	Rel +OS+RL+VSOA+E+Ren+PS
126	Rel +OS+VSA+VSOA+E+Ren+PS
127	Rel +OS+VSA+RL+E+Ren+PS
128	Rel +OS+VSA+RL+VSOA+Ren+PS
129	Rel +OS+VSA+RL+VSOA+E+PS
130	Rel +OS+VSA+RL+VSOA+E+Ren
131	Rel +OS+VSA+RL+VSOA+E+Ren+PS

No passo 1 da Tabela 11, comparou-se o modelo nulo (modelo sem covariáveis) com os modelos constituídos por uma única covariável. Os modelos comparados nessa fase foram enumerados de 1 a 14. Nesse passo, a estatística de teste utilizada foi obtida utilizando o teste da razão de verossimilhança (TRV), com uma distribuição Qui-quadrado, com um grau de liberdade dado pela diferença entre o número de covariáveis dos modelos 2 a 13 e o modelo 1. Nessa etapa, a covariável “facilidade em obter anticoncepcionais” (AO) foi retirada do processo de seleção por apresentar uma estimativa não significativa e p-valor igual a 0,7674. As covariáveis cor da pele (Cor) e programas de assistência (PA)



foram mantidas no processo de seleção, mesmo apresentando p-valor de 0,0734 e 0,1772 respectivamente, valores esses acima de 10%, mas não tão distantes.

No passo 2 da seleção das covariáveis, comparou-se o modelo completo composto pelas 12 covariáveis (Ren, Rel, Cor, MP, PS, OS, VSA, RL, PA, MC, VSOA e Dr) que foram selecionadas na etapa anterior com os modelos reduzidos compostos por 11 covariáveis (modelo completo menos uma covariável). Nessa etapa, 12 modelos reduzidos foram comparados ao modelo completo e foram enumerados de 15 (modelo completo) a 27 (modelos reduzidos). O modelo reduzido 16, que não apresenta a covariável Ren, não contribuiu significativamente, pois apresentou uma TRV de 0,192 e um p-valor de 0,6612. Já o modelo reduzido 19, que não apresentou a covariável MP, também não apresentou contribuição significativa. Sua TRV foi de 1,154 e o seu p-valor foi de 0,2827. O mesmo aconteceu com o modelo reduzido 20, que não apresentou a covariável PS, também não significativo. Sua TRV foi de 0,953 e seu p-valor foi de 0,3289. O modelo 25, que não contém a covariável MC, a exemplo dos anteriores, também não se mostrou significativo; sua TRV foi de 1,0 e seu p-valor foi de 0,3173.

Assim, as covariáveis Ren, MP, PS e MC foram excluídas da seleção das covariáveis para a próxima etapa por apresentarem contribuições não significativas. Do mesmo modo, o modelo 21, que não contém a covariável OS, também se apresentou não significativo. A covariável OS foi mantida no processo de seleção, pois a estimativa de seu p-valor 0,1103 está abaixo do limite de descarte considerado (0,25) para eliminação de covariáveis. Chegando ao final dessa etapa de seleção, as covariáveis Rel, Cor, OS, VSA, RL, PA, VSOA e Dr passarão para a próxima etapa de seleção, que é o passo 3. Essas covariáveis irão compor o modelo completo no passo 3.

No passo 3, comparou-se o modelo completo composto pelas covariáveis Rel, Cor, OS, VSA, RL, PA, VSOA e Dr aos modelos reduzidos

formados por uma covariável a menos em relação ao modelo completo. Esses modelos receberam a numeração de 28 a 36. O modelo 30, que não contém a covariável Cor, apresentou-se não significativo, pois sua TRV foi de 0,212598 e o seu p-valor de 0,6447. O mesmo aconteceu com o modelo 34 que não apresenta a covariável PA, pois sua TRV foi de 0,442444 e o seu p-valor de 0,5059. Da mesma forma, o modelo 36 que não contém a covariável Dr também não foi significativo, pois sua TRV foi de 0,000906 e o seu p-valor de 0,9760. Nesta etapa, as covariáveis Cor, PA e Dr não continuaram no processo de seleção das covariáveis, pois apresentaram contribuições não significativas com p-valor acima de 0,25.

As covariáveis restantes da etapa anterior Rel, OS, VSA, RL e VSOA compuseram o modelo completo para essa nova etapa de seleção, denominada passo 4. Seguindo-se a mesma metodologia, comparou-se o modelo completo aos modelos reduzidos compostos pela retirada de uma covariável do modelo completo. Esses modelos receberam a numeração de 37 a 42. Nota-se que até aqui todos os modelos apresentaram p-valor menor do que 25%, como mostra a Tabela 11, evidenciando que todas as covariáveis envolvidas foram significativas.

No passo 5, o modelo (43) obtido no passo 4 foi inicialmente tomado como o modelo adequado. Em seguida, tomaram-se outros modelos compostos pela adição das covariáveis que foram retiradas das etapas anteriores para verificar se são realmente não significativas, seguindo a metodologia sugerida por Collet (1994). Esses modelos receberam a numeração de 43 a 55. O modelo 44 -- contendo a covariável Mídia (M) -- mostrou-se não significativo, apresentando uma TRV de 9,248 e um p-valor de 0,3867. O mesmo aconteceu com o modelo 46, em que a covariável Escolaridade da Mãe (EM) não se mostrou significativa ao apresentar uma TRV de 1,473 e um p-valor de 0,2249. Logo, essas duas covariáveis deixam o processo de seleção. No entanto, os

modelos 45 e 47, contendo respectivamente as covariáveis Escolaridade do Pai (EP) e Escolaridade da Grávida (E) mostraram-se significativas ao nível de 0,10, pois seus p-valor foram de 0,0604 e 0,0594, porém pouco discrepantes. Para verificar essa pequena discrepância entre os modelos 45 e 47, construiu-se um modelo completo contendo as covariáveis Rel, OS, VSA, RL, VSOA, EP, E e D. Fez-se, então, a comparação com os modelos reduzidos de 49 a 55. Nota-se que os modelos 55 e 56, que contêm as covariáveis EP e E respectivamente, mostraram-se muito similares, isto é, apresentaram uma discrepância pequena entre suas TRVs (2,418292 e 2,777536) e os p-valor (0,1199 e 0,0956). Optou-se, então, por retirar do processo de seleção a covariável EP e manteve-se a covariável E, pelo fato de esta covariável estar mais ligada ao problema em estudo.

No passo 6, as covariáveis Ren, Cor, MP, PS e MC, que haviam sido excluídas temporariamente da seleção em etapas anteriores, retornaram ao processo para verificar se poderiam ser excluídas definitivamente. Para tanto, comparou-se o modelo 58 (37 do passo 4 e 43 do passo 5), que havia sido considerado o modelo inicialmente adequado, aos modelos de número 59 a 63, compostos pela adição de cada uma das citadas acima com modelo 58. Nota-se que somente o modelo 60 não se apresentou significativo, pois teve uma TRV de 0,214 e p-valor de 0,6436. Uma observação importante é que a covariável MC no modelo 63 teve um valor muito alto para a TRV (250,422) e um p-valor igual a zero. Continuando o processo, construiu-se o modelo 64 composto pelo modelo 58 com as covariáveis significativas da etapa inicial do passo 6, que foram Ren, MP, MC e PS. Nessa nova etapa, o modelo 71, que não contém a covariável MP, obteve uma TRV de 0,927 e um p-valor de 0,3365, considerado não significativo, portanto a covariável MP é excluída da seleção. Observa-se novamente que a covariável MC provocou uma alta muito grande na TRV e uma redução muito forte no p-valor. O modelo 63 comprova a observação.

No passo 7, tomou-se o modelo agora composto por 10 covariáveis, sendo 5 do modelo 58 e mais 5 selecionadas nos passos 5 e 6, que foram E, D, Ren, PS e MC. Esses modelos receberam a numeração de 74 a 84. O modelo 81, que não contém a covariável Diálogo (D), apresentou uma TRV de 0,400 e um p-valor de 0,5269, o que eliminou essa covariável do processo. Outra vez a covariável MC vem inflacionar o modelo, já o modelo 84 apresentou uma alta TRV e p-valor igual a zero.

Até o momento, considerou-se que na seleção dos modelos, os efeitos foram apenas aditivos, isto é, que nos modelos de 15 a 84 os efeitos das covariáveis foram aditivos. No passo 8, tomou-se o modelo 85, comparando-o aos modelos de 86 a 121, o que inclui um efeito multiplicativo, isto é, modelos com interação. Observou-se que uma boa parte dos modelos não foi significativa (Ex: modelos: 86, 89, 91, 92, 95, etc) e que também houve uma influência estranha desses produtos, como, por exemplo, o modelo 92, que apresentou um p-valor igual a 1. Além do mais, não se encontrou uma interpretação adequada para esses efeitos multiplicativos no contexto do problema. Portanto, não foi considerado nenhum efeito multiplicativo em nenhuma parte do trabalho.

Finalmente, no passo 9, compôs-se um modelo formado pelas covariáveis significativas em todas etapas do processo de seleção, com exceção da covariável MC que, apesar de se mostrar altamente significativa, estava inflacionando os modelos ao apresentar uma alta TRV e um p-valor igual a zero. O modelo 122 -- composto pelas covariáveis Rel, OS, VSA, RL, VSOA, Ren e PS -- foi comparado aos modelos reduzidos de número 123 a 130, para verificar se uma das covariáveis na presença de outras, deixaria de ser significativa. Entretanto, observando-se os valores das TRVs e dos p-valor, nota-se que todas foram significativas. Assim, considerou-se que o modelo 131 seria um dos modelos adequados para explicar o problema em estudo.

TABELA 11 Passos usados na seleção de covariáveis, número do modelo, logaritmo da razão de verossimilhança, estatística de teste e p-valor usados para obter o modelo paramétrico ajustado aos dados do estudo da gravidez na adolescência na microrregião de Campo Belo e Região- MG

<b>Passos</b>	<b>Número do Modelo</b>	<b>-2logL(θ)</b>	<b>Estatística de Teste (TRV)</b>	<b>p-valor</b>
Passo1	1	7664,989	-	-
	2	7653,351	11,638	0,0005
	3	7655,246	9,743	0,0016
	4	7661,765	3,224	0,0734
	5	7654,433	10,556	0,0014
	6	7644,281	20,708	<0,0001
	7	7613,574	51,415	<0,0001
	8	7570,962	94,027	<0,0001
	9	7604,956	60,033	<0,0001
	10	7663,156	1,833	0,1772
	11	7367,900	297,089	<0,0001
	12	7605,289	59,700	<0,0001
	13	7659,694	5,295	0,0234
	14	7664,901	0,088	0,7674
Passo2	15	7203,835	-	-
	16	7203,643	0,192	0,6612
	17	7212,882	9,047	0,0026
	18	7211,490	7,655	0,0057
	19	7204,989	1,154	0,2827
	20	7204,788	0,953	0,3289
	21	7206,384	2,549	0,1103
	22	7209,635	5,800	0,0160
	23	7244,275	40,440	<,0001
	24	7219,756	15,921	<,0001
	25	7204,835	1,000	0,3173
	26	7443,297	239,462	0,000
	27	7218,832	14,997	0,00011

(Continua...)

“TABELA 11, Cont.”

Passo3	28	7474,229	-	-
	29	7480,907	6,677614	0,0098
	30	7474,442	0,212598	0,6447
	31	7506,595	32,36631	<0,001
	32	7512,956	38,72703	<0,001
	33	7509,800	35,57138	<0,001
	34	7474,671	0,442444	0,5059
	35	7482,788	8,558844	0,0035
	36	7474,230	0,000906	0,9760
Passo4	37	7484,138	-	-
	38	7474,890	9,248	0,0023
	39	7481,528	2,610	0,1061
	40	7507,910	23,772	<,0001
	41	7516,087	31,949	<,0001
	42	7510,638	26,500	<,0001
Passo5	43	7474,890	-	-
	44	7474,141	0,749	0,3867
	45	7471,363	3,527	0,0604
	46	7473,417	1,473	0,2249
	47	7471,337	3,553	0,0594
	48	7471,731	3,159	0,0755
	49	7464,872	-	-
	50	7471,333	6,460762	0,0110
	51	7502,318	37,44591	<0,001
	52	7505,698	40,82611	<0,001
	53	7503,198	38,32576	<0,001
	54	7473,424	8,551958	0,0035
	55	7467,29	2,418292	0,1199
	56	7467,65	2,777536	0,0956
	57	7469,22	4,348138	0,0370

(Continua...)

“TABELA 11, Cont.”

Passo6	58	7474,890	-	-
	59	7464,676	10,214	0,0014
	60	7474,676	0,214	0,6436
	61	7459,677	15,213	<,0001
	62	7463,497	11,393	0,0007
	63	7224,468	250,422	0,000
	64	7209,294	-	-
	65	7217,082	7,787664	0,0053
	66	7215,098	5,804146	0,0156
	67	7254,87	45,57561	<0,0001
	68	7224,504	15,21044	<0,0001
	69	7222,819	13,52535	0,0003
	70	7217,275	7,981206	0,0047
	71	7210,221	0,927006	0,3365
72	7446,84	237,5462	0,0000	
Passo7	73	7212,243	2,948614	0,0859
	74	7201,487	-	-
	75	7209,532	8,044826	0,0046
	76	7210,593	9,106034	0,0025
	77	7244,983	43,49566	<0,001
	78	7217,189	15,7015	<0,001
	79	7212,629	11,14243	0,0008
	80	7210,115	8,627936	0,0033
	81	7201,887	0,40032	0,5269
	82	7211,821	10,33438	0,0013
	83	7207,358	5,871338	0,0154
	84	7445,213	243,7258	0,0000
	85	7210,221	-	-
	86	7209,706	0,515	0,4730
	87	7206,643	3,578	0,0585
	88	7203,632	6,589	0,0103
	89	7207,746	2,475	0,1157
	90	7205,129	5,092	0,0240
	91	7210,203	0,018	0,8933
	92	7210,221	0,000	1,0000
	93	7210,047	0,174	0,6766
	94	7192,958	17,263	<,0001
	95	7210,199	0,022	0,8821
	96	7208,065	2,156	0,1420
	97	7206,587	3,634	0,0566
	98	7203,188	7,033	0,0080

(Continua...)

“TABELA 11, Cont.”

Passo 7, “Cont.”	99	7182,725	27,496	<,0001
	100	7195,233	14,988	0,0001
	101	7208,292	1,929	0,1649
	102	7210,019	0,202	0,6531
	103	7198,147	12,074	0,0005
	104	7207,379	2,842	0,0918
	105	7179,710	30,511	<,0001
	106	7181,220	29,001	<,0001
	107	7208,180	2,041	0,1531
	108	7185,779	24,442	<,0001
	109	7203,331	6,890	0,0087
	110	7195,572	14,649	0,0001
	111	7202,439	7,782	0,0053
	112	7202,925	7,296	0,0069
	113	7209,638	0,583	0,4451
	114	7203,408	6,813	0,0090
	115	7192,751	17,470	<,0001
	116	7206,731	3,490	0,0617
	117	7202,591	7,630	0,0057
	118	7208,941	1,280	0,2579
	119	7207,089	3,132	0,0768
120	7210,176	0,045	0,8320	
121	7209,969	0,252	0,6157	
Passo 9	122	7447,657	-	-
	123	7453,468	5,8108	0,0159
	124	7480,57	32,91332	<0,001
	125	7486,319	38,66232	<0,001
	126	7484,017	36,36038	<0,001
	127	7455,246	7,588576	0,0059
	128	7453,892	6,235364	0,0125
	129	7459,63	11,97344	0,0005
	130	7458,743	11,08615	0,0009
	Passo 10	131	7447,657	-

### 4.3 Adequação do modelo

Escolhido o conjunto de covariáveis, passou-se a discriminar entre os modelos de regressão Exponencial, Weibull e Log-Normal, que são casos particulares do modelo Gama Generalizado, a fim de se determinar qual será utilizado para a interpretação das estimativas dos parâmetros das covariáveis.



Nota-se pela Tabela 12, que os três modelos encaixados foram significativos, com  $p < 0,0001$ . No entanto, o modelo Log-Normal apresentou o menor valor da TRV (176,19), sendo então o modelo escolhido.

TABELA 12 Modelos, Logaritmo da função  $L(\theta)$ , resultados da TRV e p-valor usados para obtenção do modelo paramétrico ajustado aos dados do estudo da gravidez na adolescência na microrregião de Campo Belo-MG

<i>Modelo</i>	<i>Log(L<math>\theta</math>)</i>	<i>TRV</i>	<i>p-valor</i>
Gama	-3723,829	-	
Generalizado	-3723,829	-	
Exponencial	-4042,669	$-2[-4042,669 - (-3723,829)] = 637,68$	0
Weibull	-4027,732	$-2[-4027,732 - (-3723,829)] = 601,81$	0
Log-normal	-3811,923	$-2[-3811,923 - (-3723,829)] = 176,19$	0

Para confirmar essa escolha, foram feitos os gráficos dos resíduos padronizados versus a curva de sobrevivência estimada para o modelo Exponencial, Weibull e Log-Normal. Esses gráficos são compostos por duas linhas, sendo uma pontilhada (representando a curva de sobrevivência dos resíduos da função) e a outra cheia (representando a curva de sobrevivência por Kaplan & Méier (1958)). O gráfico em que as linhas estiverem mais próximas, define o modelo mais ajustado, embora o ideal fosse a superposição delas. Sendo assim, nota-se que, dos três gráficos da figura 2, o gráfico 3, que é o do modelo Log-Normal, ajustou-se melhor aos dados.

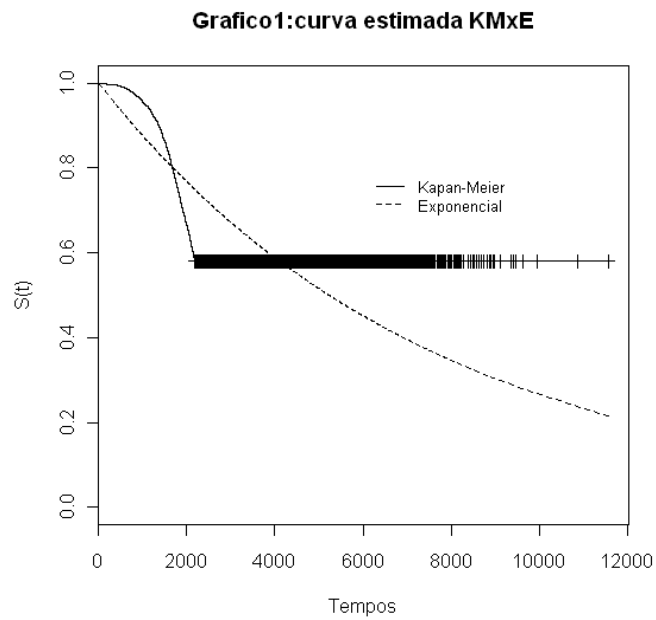
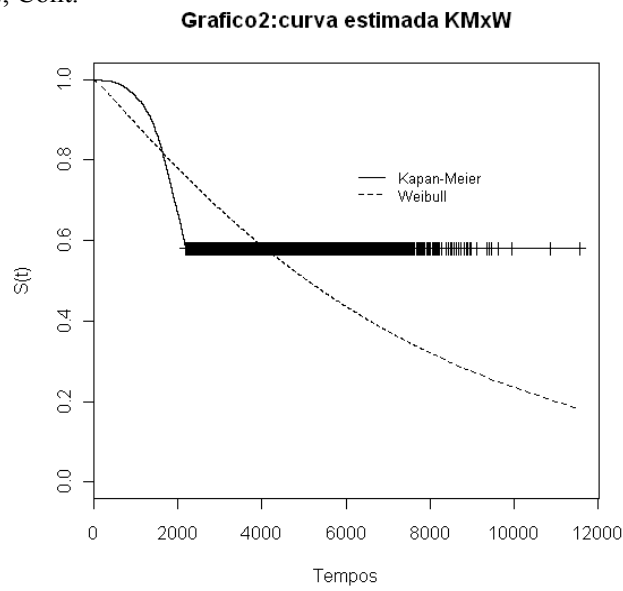


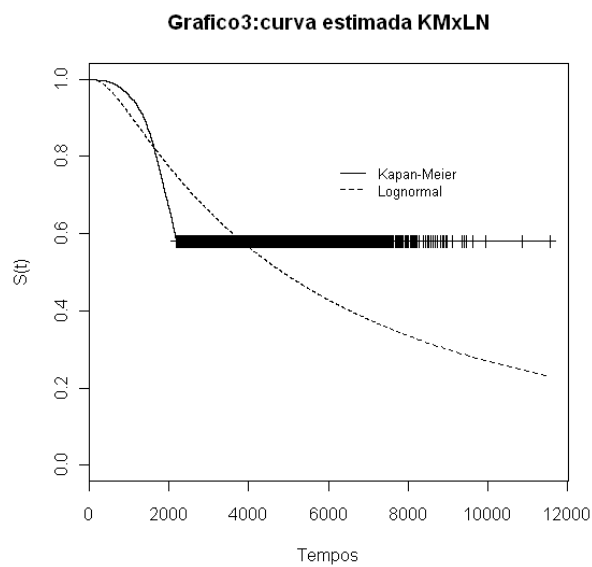
FIGURA 2 Representações gráficas das análises dos resíduos de Cox-Snell para os modelos log-normal, exponencial e weibull ajustados aos dados. (Continua...)

“FIGURA 2, Cont.”



(Continua...)

“FIGURA 2, Cont.”



Fez-se também a representação dos resíduos padronizados estimados por Kaplan-Meier versus o gráfico da curva sobrevivência estimada pelo modelo Log-Normal. Nota-se que uma boa parte dos pontos ficou sobre uma reta (Figura 3). Também foi feita (Figura 4) a representação dos resíduos de Cox-Snell estimados por Kaplan & Meeir (1958) versus a sobrevivência dos mesmos estimados pelo modelo Exponencial padrão. Também, observa-se que uma boa parte dos pontos está sobre uma reta. Isso garante que o modelo Log-Normal apresenta-se ajustado.

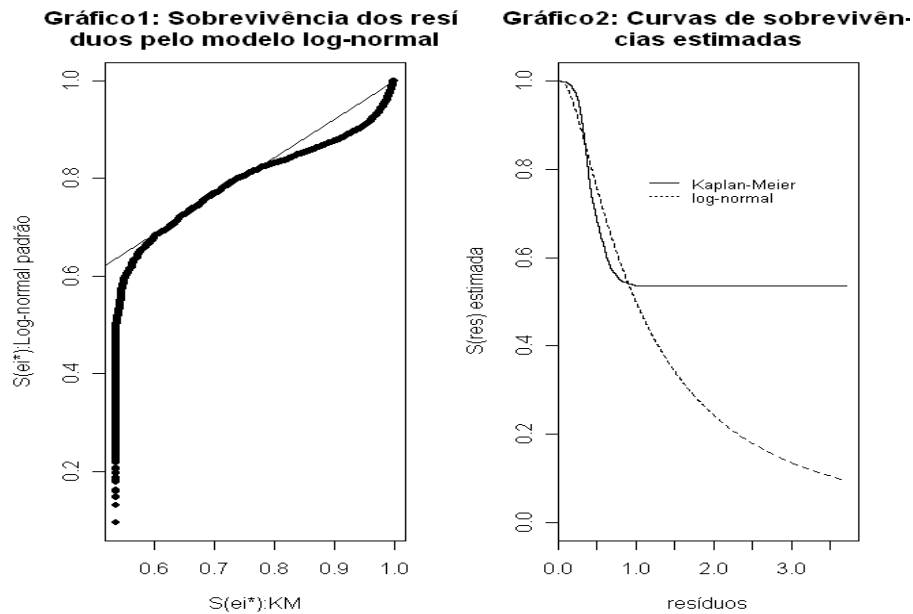


FIGURA 3 Representações gráficas da sobrevivência dos resíduos estimados por K-M e pelo modelo Log-Normal e da curva sobrevivência estimada.

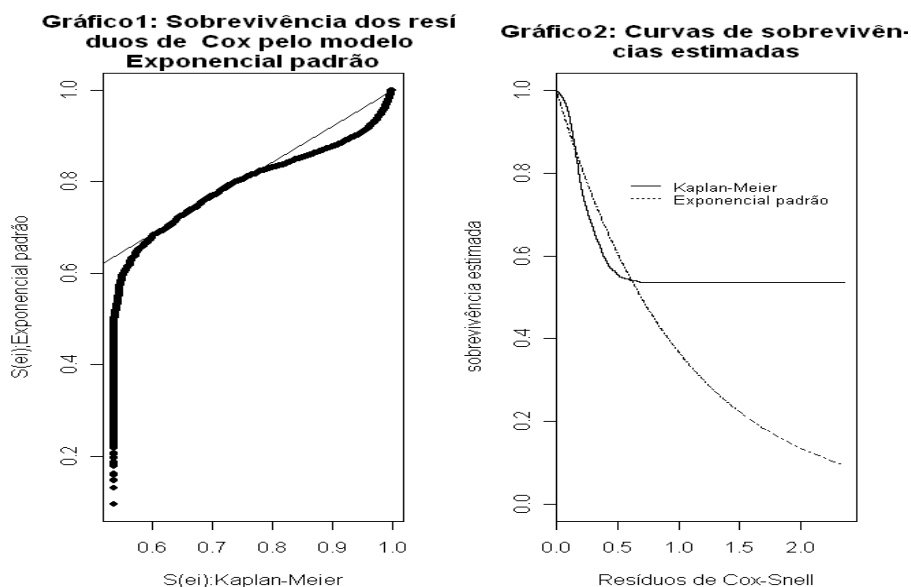


FIGURA 4 Representações gráficas das curvas de sobrevivência dos resíduos estimados por K-M e pelo modelo Exponencial padrão e da curva sobrevivência estimada.

#### 4.4 Interpretação das estimativas

A Tabela 13 mostra as estimativas dos parâmetros do modelo de regressão Log-Normal ajustado aos dados da gravidez na adolescência.

TABELA 13 Estimativas dos parâmetros, erro-padrão, valor de Z e p-valor do modelo de regressão Log- Normal ajustados aos dados do estudo da gravidez na adolescência na microrregião de Campo Belo-MG

Parâmetro	Estimativa	Erro-padrão	Valor de $Z_c$	$P <  Z_c $
Intercepto	7,891	0,0969	81,41	0,0000
Rel	-0,124	0,0526	-2,37	0,0179
OS	0,209	0,0534	3,22	0,0013
VSA	-0,398	0,0502	8,08	<0,0001

(Continua...)

“TABELA 13, Cont.”

<i>RL</i>	-0,328	0,0690	-4,57	<0,0001
VSOA	-0,169	0,0491	3,47	0,0005
E	-0,210	0,0293	-2,98	0,00289
Ren	0,214	0,0344	4,41	<0,001
PS	-0,142	0,0529	2,70	0,00704

O modelo Log-Normal<sup>4</sup> pode ser escrito da seguinte maneira:

$$\log(T) = 7,891 - 0,124 * Rel + 0,209 * OS - 0,398 * VSA - 0,328 * RL - 0,169 * VSOA - 0,210 * E + 0,214 * Ren - 0,142 * PS + \sigma Z$$

Em que Z tem uma distribuição normal padrão.

A interpretação das estimativas do tempo mediano não é direta, visto que o tempo está logado. Então para interpretar qualquer uma das estimativas do tempo mediano tem-se que fazer a operação inversa, isto é, tomar a exponencial da estimativa.

Por exemplo, considere dois indivíduos da covariável Religião (Rel),

tem-se que  $\frac{t_{0,5}(x=1, \hat{\beta})}{t_{0,5}(x=0, \hat{\beta})} = e^{\hat{\beta}} = e^{-0,124} = 1,132$ , isto significa que as mulheres das religiões evangélicas (x=0) apresentam um tempo mediano de engravidar na adolescência 1,132 vezes maior do que as mulheres da religião católica (x=1), mantidas fixas as outras covariáveis<sup>1</sup>. Para facilitar o entendimento, pode-se considerar que as mulheres das religiões evangélicas apresentam um aumento de 13% no tempo mediano em relação às mulheres da religião católica, isto é, a

<sup>4</sup> Pode haver outros modelos envolvendo outras covariáveis do estudo, no entanto, houve preferência por essas pelo fato de que o pesquisador, baseado em suas observações, julgou ser este modelo o de maior poder de explicação dos dados da gravidez na adolescência na microrregião de Campo Belo-MG.

possibilidade de elas engravidarem na adolescência é 13% menor em relação às outras.

Baseando-se em interpretações como a anterior, verifica-se que:

- a possibilidade de as mulheres evangélicas engravidarem na adolescência é 13% menor do que entre as católicas. Este fato contraria resultados de Aquino et al. (2003), ao afirmar que as mulheres da religião católica apresentam menor experiência sexual antes dos vinte anos e confirma os resultados de Carvalho et al. (1999) e Guimarães (2001), cujo estudo indica que as religiões evangélicas apresentam maior rigidez no que diz respeito à prática sexual;
- a possibilidade de engravidar na adolescência entre as mulheres que tiveram algum tipo de orientação sexual foi 18,8% menor do que as que não tiveram nenhum tipo de orientação;
- a possibilidade de não engravidar na adolescência entre as mulheres que não tiveram uma vida sexual ativa foi 33% menor em relação às que tiveram;
- a possibilidade de engravidar na adolescência entre as mulheres que tiveram conhecimento sobre a lei de paternidade foi 38,8% maior em relação às aquelas que desconheciam tal lei;
- a possibilidade de engravidar na adolescência entre as mulheres que não tiveram uma vida social ativa foi 15,54% menor em relação às aquelas que tiveram.
- a possibilidade de engravidar na adolescência entre as mulheres que apresentaram escolaridade até 4ª série foi 23,3% menor em relação às aquelas com escolaridade inferior a nível. Este fato está de acordo

com os resultados de Aquino (2003), Baraldi et al (2007) e Ximenes Neto et al. (2007).

- a possibilidade de engravidar na adolescência entre as mulheres com renda inferior ou igual a um salário mínimo foi 19,3% maior em relação àquelas com salário superior ao mínimo; resultados compatíveis com os de Aquino et al. (2003) e Borges & Schor (2005).
- a possibilidade de engravidar na adolescência entre as mulheres que tiveram os pais separados foi 13,2% maior de do que aquelas cujos pais não se separaram, como mostra Young (1991).

#### 4.5 Análise no modelo de regressão de Cox

Nesta seção, foi usada a mesma estratégia de seleção de covariáveis utilizadas na seleção do modelo paramétrico (seção anterior).

A Tabela 14 mostra os modelos usados nos na seleção do modelo de riscos proporcionais de Cox.

TABELA 14 Número do modelo e descrição do modelo usado para obter o modelo de riscos proporcionais Cox ajustado aos dados do estudo da gravidez na adolescência na microrregião de Campo Belo-MG

Número do Modelo	Modelo
1	Nulo
2	Ren
3	Rel
4	Cor
5	E
6	MP
7	PS

(Continua...)



“TABELA 14, Cont.”

8	OS
9	VSA
10	RL
11	PA
12	MC
13	VSOA
14	Dr
15	Ren+E+MP+PS+OS+VSA+RL+PA+MC+VSOA+Dr
16	E+MP+PS+OS+VSA+RL+PA+MC+VSOA+Dr
17	Ren+MP+PS+OS+VSA+RL+PA+MC+VSOA+Dr
18	Ren+E+PS+OS+VSA+RL+PA+MC+VSOA+Dr
19	Ren+E+MP+OS+VSA+RL+PA+MC+VSOA+Dr
20	Ren+E+MP+PS+VSA+RL+PA+MC+VSOA+Dr
21	Ren+E+MP+PS+OS+RL+PA+MC+VSOA+Dr
22	Ren+E+MP+PS+OS+VSA+PA+MC+VSOA+Dr
23	Ren+E+MP+PS+OS+VSA+RL+MC+VSOA+Dr
24	Ren+E+MP+PS+OS+VSA+RL+PA+VSOA+Dr
25	Ren+E+MP+PS+OS+VSA+RL+PA+MC+Dr
26	Ren+E+MP+PS+OS+VSA+RL+PA+MC+VSOA
27	Ren +E+ PS+OS+RL+ PA+MC+VSOA
28	E+ PS+OS+RL+ PA+MC+VSOA
29	Ren+ PS+OS+RL+ PA+MC+VSOA
30	Ren +E+OS+RL+ PA+MC+VSOA
31	Ren +E+ PS+RL+ PA+MC+VSOA
32	Ren +E+ PS+OS+ PA+MC+VSOA
33	Ren +E+ PS+OS+RL+MC+VSOA
34	Ren +E+ PS+OS+RL+ PA+VSOA
35	Ren +E+ PS+OS+RL+ PA+MC
36	Ren +E+ PS+PA+MC+VSOA
37	E+ PS+PA+MC+VSOA
38	Ren+ PS+PA+MC+VSOA
39	Ren +E+PA+MC+VSOA
40	Ren +E+ PS+MC+VSOA
41	Ren +E+ PS+PA+VSOA
42	Ren +E+ PS+PA+MC
43	Ren +E+ PS+PA+MC+VSOA
44	Ren +E+ PS+PA+MC+VSOA+Rel
45	Ren +E+ PS+PA+MC+VSOA+Cor
46	Ren +E+ PS+PA+MC+VSOA+EP
47	Ren +E+ PS+PA+MC+VSOA+EM
48	Ren +E+ PS+PA+MC+VSOA+Dr
49	Ren +E+ PS+PA+MC+VSOA+M

(Continua...)

“TABELA 14, Cont.”

50	Ren +E+ PS+PA+MC+VSOA+VSA
51	Ren +E+ PS+PA+MC+VSOA+RL
52	Ren +E+ PS+PA+MC+VSOA+D
53	Ren +E+ PS+PA+MC+VSOA+OS
54	Ren +E+ PS+PA+MC+VSOA+PA
55	Ren +E+ PS+PA+MC+VSOA+AO
56	Ren +E+ PS+RL+VSOA+Rel+VSA+OS+D
57	E+ PS+RL+VSOA+Rel+VSA+OS+D
58	Ren+ PS+RL+VSOA+Rel+VSA+OS+D
59	Ren +E+RL+VSOA+Rel+VSA+OS+D
60	Ren +E+ PS+VSOA+Rel+VSA+OS+D
61	Ren +E+ PS+RL+Rel+VSA+OS+D
62	Ren +E+ PS+RL+VSOA+VSA+OS+D
63	Ren +E+ PS+RL+VSOA+Rel+OS+D
64	Ren +E+ PS+RL+VSOA+Rel+VSA+D
65	Ren +E+ PS+RL+VSOA+Rel+VSA+OS
66	Ren +E+ PS+RL+VSOA+Rel+VSA+OS
67	E+ PS+RL+VSOA+Rel+VSA+OS
68	Ren + PS+RL+VSOA+Rel+VSA+OS
69	Ren +E+RL+VSOA+Rel+VSA+OS
70	Ren +E+ PS+VSOA+Rel+VSA+OS
71	Ren +E+ PS+RL+Rel+VSA+OS
72	Ren +E+ PS+RL+VSOAI+VSA+OS
73	Ren +E+ PS+RL+VSOA+Rel+OS
74	Ren +E+ PS+RL+VSOA+Rel+VSA
75	Ren +E+ PS+RL+VSOA+Rel+VSA+OS

Nota-se, observando o passo 1 da Tabela 15, que as covariáveis Rel e Cor (modelos 3 e 4, respectivamente), por apresentarem p-valor iguais a 0,8060 e 0,6147 (superiores a 0,10), serão retiradas momentaneamente do processo de seleção.

Na etapa seguinte, passo 2, tomou-se o modelo completo formado pelas covariáveis selecionadas na etapa anterior, passo 1, comparando-o aos modelos reduzidos formados por uma covariável a menos do que no modelo completo. Foram comparados 11 modelos, enumerados de 15 (modelo completo) a 26. Nesse passo, a estatística de teste foi obtida utilizando o teste da razão de verossimilhança (TRV) com uma distribuição Qui-quadrado com um grau de

liberdade dado pela diferença entre o número de covariáveis dos modelos. Esse procedimento foi usado em todos os passos do processo de seleção de covariáveis. Nota-se que os modelos 18, 22 e 26 apresentaram p-valor de 0,5598, 0,16808 e 0,3222 respectivamente e que esses valores são superiores a 0,10, que foi o nível de significância tomado com referência. Portanto, as covariáveis MP e Dr, ausentes nos modelos citados, foram retiradas temporariamente da seleção. A covariável RL, por apresentar um p-valor não tão distante de 0,10, comparado às demais, foi mantida na seleção, passando assim para a próxima etapa.

No passo 3, tomou-se o modelo completo formado pelas covariáveis selecionadas no passo 2, modelo 27, comparando-o aos modelos reduzidos enumerados de 28 a 35. Nota-se que os modelos 31 e 32 apresentaram p-valor de 0,50712 e 0,22693 respectivamente. As covariáveis Os e RL ausentes nesses modelos foram também temporariamente retiradas do processo de seleção.

No passo 4, tomou-se o modelo completo (modelo 36) formado pelas covariáveis Ren, E, PS, PA MC e VSOA selecionadas na etapa anterior, comparando-o aos modelos reduzidos enumerados de 37 a 42. Nota-se que todos os modelos foram significativos ao nível de significância 0,10, pois todos apresentaram p-valor inferior a esse nível. Então, o modelo 36 foi inicialmente considerado como o modelo adequado.

No passo 5, comparou-se o modelo 43 (36 do passo 4) com os modelos de 44 a 56. Esses modelos foram construídos adicionando-se as covariáveis, que antes haviam sido retiradas do processo de seleção nos passos anteriores, ao modelo 43. Nota-se que os modelos 45, 46, 47, 48, 49, 55 e 56 apresentaram p-valor superiores a 0,10. Logo, as covariáveis Cor, EP, EM, Dr, M, PA e AO presentes nesses modelos foram retiradas definitivamente do processo. A covariável RL, mesmo apresentando p-valor superior a 0,10, foi mantida. Já a

covariável MC, pela mesma razão verificada na modelagem paramétrica foi retirada do processo.

No passo 6, tomou-se o modelo completo (modelo 56) formado pelas covariáveis selecionadas no passo anterior, comparando-o aos modelos reduzidos de 57 a 65. Nota-se que apenas o modelo 65 apresentou um p-valor superior a 0,10. Isto fez com que a covariável D, ausente nesse modelo, deixasse definitivamente a seleção.

No passo 7, construiu-se o modelo completo (modelo 66) formado pelas covariáveis Ren, E, PS, RL, VSOA, Rel, VSA e OS, comparando-o aos modelos reduzidos de 67 a 74. Nota-se que todos eles apresentaram p-valor inferior a 0,10, mostrando com isso que todas as covariáveis envolvidas foram significativas. Então, o modelo 75 -- composto pelas covariáveis citadas acima -- foi considerado o modelo definitivo.

TABELA 15 Passos usados na seleção de covariáveis, número do modelo, modelo, logaritmo da razão de verossimilhança, estatística de teste e p-valor, usados para obter o modelo de riscos proporcionais Cox ajustado aos dados do estudo da gravidez na adolescência na microrregião de Campo Belo-MG

<b>Passos</b>	<b>Número do Modelo</b>	<b>-2logL(<math>\theta</math>)</b>	<b>Estatística de Teste (TRV)</b>	<b>p-valor</b>
Passo 1	1	25586,18	-	-
	2	25581,36	4,82	0,0281
	3	25586,12	0,06	0,806
	4	25585,98	0,2	0,6547
	5	25573,22	12,96	0,0003
	6	25582,86	3,32	0,0684
	7	25557,98	28,2	<0,0001
	8	25560,74	25,44	<0,0001
	9	25450,98	135,2	0,0000
	10	25567,68	18,5	<0,0001
	11	25570,64	15,54	<0,0001
	12	25349,32	236,86	0,0000
	13	25513,88	72,3	0,0000
	14	25576,7	9,48	0,0021

(Continua...)

“TABELA 15, Cont.”

Passo2	15	25202,34	-	-
	16	25213,68	11,34	0,0008
	17	25218,16	15,82	<0,0001
	18	25202,68	0,34	0,5598
	19	25206,44	4,1	0,04288
	20	25205,56	3,22	0,07274
	21	25251,26	48,92	<0,0001
	22	25204,24	1,9	0,16808
	23	25206,38	4,04	0,04443
	24	25377,22	174,88	0,0000
	25	25209,04	6,7	0,00964
26	25203,32	0,98	0,32220	
Passo3	27	25251,42	-	-
	28	25260,06	8,64	0,00329
	29	25272,14	20,72	<0,0001
	30	25259,08	7,66	0,0056
	31	25251,86	0,44	0,50712
	32	25252,88	1,46	0,22693
	33	25264,34	12,92	0,00033
	34	25436,62	185,2	0,0000
	35	25280,68	29,26	0,0001
Passo4	36	25253,22	-	-
	37	25262	8,78	0,00305
	38	25275,6	22,38	<0,001
	39	25260,76	7,54	0,0060
	40	25265,44	12,22	0,00047
	41	25471,52	218,3	0,0000
	42	25282,3	29,08	<0,001
Passo5	43	25253,22	-	-
	44	25262	8,78	0,0030
	45	25252,6	0,62	0,4310
	46	25252,320	0,9	0,3428
	47	25253,16	0,06	0,8065
	48	25253,06	0,16	0,6892
	49	25253,16	0,06	0,8065
	50	25208,66	44,56	<0,001
	51	25251,86	1,36	0,2435
	52	25377,22	124	0,0000
	53	25252,88	0,34	0,5598
	54	25253,220	0	1,0000
	55	25253,220	0	1

(Continua...)

“TABELA 15, Cont.”

Passo6	56	25372,96	-	-
	57	25388,72	15,76	<0,001
	58	25390,56	17,6	<0,001
	59	25376,88	3,92	0,0477
	60	25391,5	18,54	<0,001
	61	25387,14	14,18	<0,001
	62	25375,96	3	0,0833
	63	25437,34	64,38	<0,001
	64	25376,9	3,94	0,0471
Passo7	65	25375,06	2,1	0,1473
	66	25375,06	-	-
	67	25392,54	17,48	<0,0001
	68	25391,6	16,54	<0,0001
	69	25379,32	4,26	0,0390
	70	25393,68	18,62	0,0001
	71	25388,76	13,7	0,00021
	72	25378,1	3,04	0,08123
	73	25437,86	62,8	<0,0001
Passo 8	74	25379,04	3,98	0,04604
	75	25375,06	-	-

#### 4.6 Interpretação das estimativas

A Tabela 16 mostra as estimativas dos parâmetros do modelo de riscos proporcionais de Cox ajustado aos dados da gravidez na adolescência.

TABELA 16 Estimativas dos parâmetros ( $\beta$ ), seus respectivos erros-padrão (EP) e p-valor, razão de risco (RR) e intervalo de confiança da razão de risco relativos ao ajuste do modelo de riscos proporcionais de Cox ajustado aos dados da gravidez na adolescência na microrregião de Campo Belo-MG

Covariável	Estimativa( $\beta$ )	EP	p-valor	RR( $e^{\beta}$ )	IC <sub>95%</sub> (RR)
E	0,220	0,0544	<0,001	1,246	[1,120;1,386]
PS	-0,129	0,0617	0,037	0,879	[0,779;0,992]
VSA	-0,475	0,0592	<0,001	0,622	[0,554;0,699]

(Continua...)

“TABELA 16, Cont.”

OS	-0,148	0,0732	0,03	0,863	[0,747;0,996]
VSOA	-0,213	0,0572	0,002	0,808	[0,723;0,904]
RL	0,358	0,0797	<0,001	1,431	[1,224;1,673]
Rel	0,110	0,0637	0,084	1,116	[0,985;1,265]
Ren	-0,256	0,0626	<0,001	0,774	[0,685;0,875]

O modelo pode ser escrito da seguinte maneira:

$$\alpha_{ij} = \alpha_{oj}(t) \exp(0,220 * E - 0,129 * PS - 0,475 * VSA - 0,148 * OS - -0,213 * VSOA + 0,358 * RL + 0,110 * Rel - 0,256 * Ren)$$

Em que  $\alpha_{oj}$  é a função de risco basal.

A obtenção da razão de risco para os estratos de uma determinada covariável é feita a exemplo da covariável religião, considerando as demais covariáveis fixas. Tem-se então:

$$\frac{\alpha(x=1)}{\alpha(x=0)} = e^{0,110} = 1,116$$

isto é feito para cada covariável.

A interpretação de cada covariável é feita da seguinte maneira:

- o risco de engravidar na adolescência entre as mulheres da religião católica apresentou-se 1,116 vezes maior do que entre aquelas de religiões evangélicas. Além disso, pode-se dizer com 95% de confiança que esse risco varia entre 0,985 e 1,265;
- as mulheres filhas de pais não separados apresentaram um risco de engravidar na adolescência 0,879 vezes maior do que as filhas cujos

pais que se separaram. Além disso, pode-se dizer com 95% de confiança que esse risco varia entre 0,779 e 0,992;

- as mulheres que não tiveram vida sexual ativa apresentaram um risco 0,622 vezes maior de engravidar na adolescência do que aquelas que tiveram. Além disso, pode-se dizer com 95% de confiança que esse risco varia entre 0,554 e 0,622;
- as mulheres que tiveram orientação sexual apresentaram um risco 0,863 vezes maior de engravidar na adolescência do que aquelas que não tiveram qualquer orientação. Além disso, pode-se dizer com 95% de confiança que esse risco varia entre 0,747 e 0,996;
- as mulheres que não tiveram vida social ativa apresentaram 0,808 vezes maior o risco de engravidar na adolescência do que aquelas que tiveram tal comportamento. Além disso, pode-se dizer com 95% de confiança que esse risco varia entre 0,723 e 0,904;
- as mulheres que tiveram conhecimento sobre as leis de paternidade apresentaram 1,431 vezes maior o risco de engravidar na adolescência do que as que não apresentaram ter tal conhecimento. Além disso, pode-se dizer com 95% de confiança que esse risco varia entre 1,224 e 1,673;
- as mulheres que tinham renda superior ou igual a um salário mínimo apresentaram 0,774 vezes maior o risco de engravidar na adolescência do que as com renda inferior a um salário mínimo. Além disso, pode-se dizer com 95% de confiança que esse risco varia entre 0,685 e 0,875;
- as mulheres com escolaridade inferior ou igual à 4ª série primária ou ao 5º ano apresentaram um risco de engravidar na adolescência 1,246 vezes maior do que aquelas com escolaridade superior ao ponto de



corde. Além disso, pode-se dizer com 95% de confiança que esse risco varia entre 1,120 e 1,386.

#### **4.7 Seleção, avaliação e adequação do modelo de Cox - riscos proporcionais**

Observando-se a Figura 5, nota-se que as curvas da religião tendem a cruzar-se bem no extremo, mostrando que existe uma possibilidade de violação da proporcionalidade de risco, o que pode ser confirmado pela Figura 8, visto que a linha (*lowsess*) no início e no final estão muito próximas dos limites inferior e superior do intervalo de confiança. Além disso, a tabela 9 mostra que a covariável Religião apresenta p-valor de 0,024, rejeitando a hipótese de proporcionalidade de risco ao nível de significância de 0,05.

Já o gráfico 2 da Figura 5 para a covariável Renda, leva a pensar que a proporcionalidade do risco foi violada, visto que as duas curvas se confundem. No entanto, a Figura 7 mostra o contrário, isto é, que os riscos são proporcionais, pois a linha *lowsess* está dentro das linhas do intervalo de confiança, o que é confirmado pela Tabela 9, visto que a covariável Ren apresenta p-valor de 0,378, ou seja, não há indícios para suspeitar da não proporcionalidade de risco.

O gráfico 3 da Figura 5 mostra indícios de que a proporcionalidade do risco para a covariável Pais Separados (PS) é violada, pois as linhas cruzam-se no início. No entanto, pelo gráfico da Figura 13, nota-se que não houve violação da pressuposição de riscos proporcionais, visto que a linha *lowsess* permanece dentro das linhas do intervalo de confiança e pela tabela 8, também nota-se que não houve a violação da pressuposição de risco, visto que a covariável PS apresenta p-valor de 0,0879.

Situação semelhante é vista no gráfico 4 da Figura 5, em que a covariável VSA também apresenta indícios de violação da proporcionalidade dos riscos, apesar de que o cruzamento das linhas ocorre muito no início e de

maneira sutil, mantendo-se um afastamento constante até no final. No entanto, a violação da suposição de riscos proporcionais é negada pelo gráfico da Figura 12, pela mesma razão anterior, e pela Tabela 9, já que a covariável VSA apresenta p-valor de 0,359, rejeitando a não proporcionalidade dos riscos.

Já nos gráficos 5 e 6 da Figura 6, nota-se que suas linhas estão bem distantes no início e vão ficando cada vez mais próximas, mostrando que há motivos para suspeitar que a violação da suposição de riscos proporcionais das covariáveis OS e RL pode ocorrer, visto que existe uma tendência de eles se cruzarem no final o que é confirmado pelos gráficos das Figuras 12 e 10 respectivamente, pois a linha lowess deixa as linhas do intervalo de confiança; isso é confirmado pela Tabela 9, uma vez que as covariáveis OS e RL apresentam p-valor de 0,0013 e  $<0,0001$  respectivamente, rejeitando-se, assim, a hipótese da proporcionalidade dos riscos.

O gráfico 7 da Figura 6 mostra que pode não ocorrer a violação da proporcionalidade de risco para a covariável VSOA, apesar de que no início está ocorrendo uma superposição muito discreta das linhas, o que é confirmado pela Figura 10 e pela Tabela 9 em que a covariável VSOA apresenta p-valor de 0,353, rejeitando-se a não proporcionalidade do risco. E por fim, o gráfico 8 da Figura 6 mostra que existem motivos para suspeitar da violação da proporcionalidade do risco para covariável E, o que é confirmado pelo gráfico da Figura 14, em que a linha lowess está muito próxima dos limites do intervalo de confiança, e também pela Tabela 9, em que o p-valor da covariável E é de 0,0315, rejeitando com isso a proporcionalidade dos riscos.

Pelo que foi exposto acima, as covariáveis E, OS, RL e Rel mostraram violar o princípio da proporcionalidade de riscos, sendo necessário fazer um estudo do modelo de Cox estratificado para essas covariáveis, o que será feito na próxima seção.

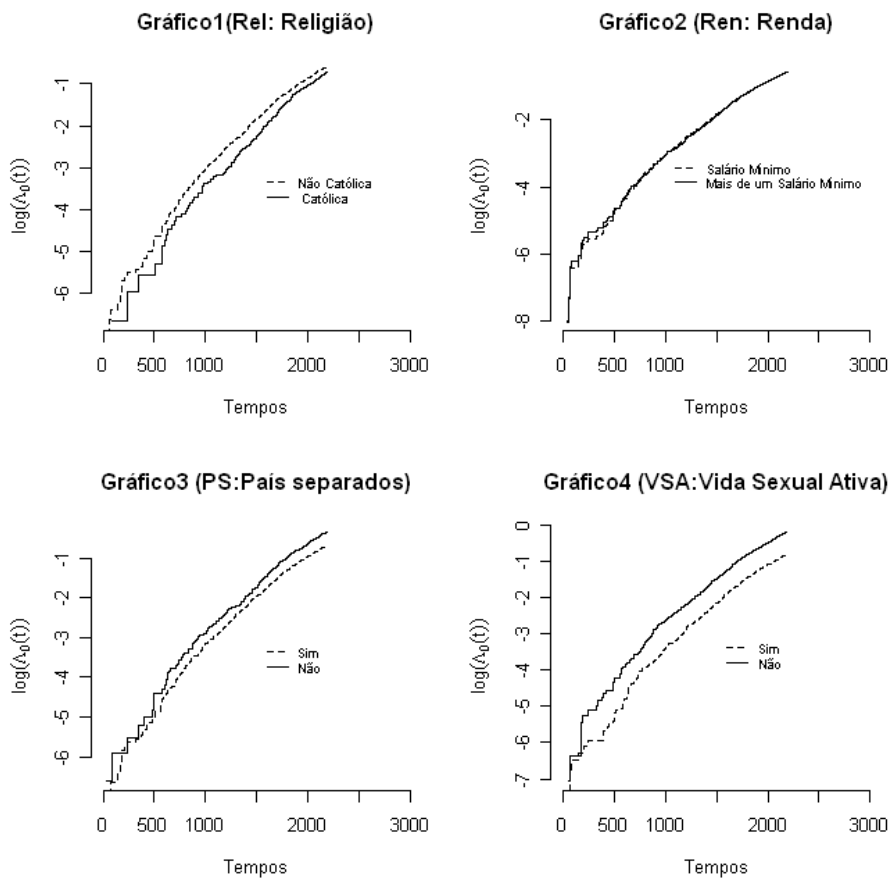


FIGURA 5 Representações gráficas do logaritmo do risco  $\text{Log}(\hat{\Delta}_{0j}(t))$  em relação ao tempo em dias após a gravidez para as covariáveis religião (Rel), renda (Ren), programas sociais (OS) e vida social ativa (VSA).

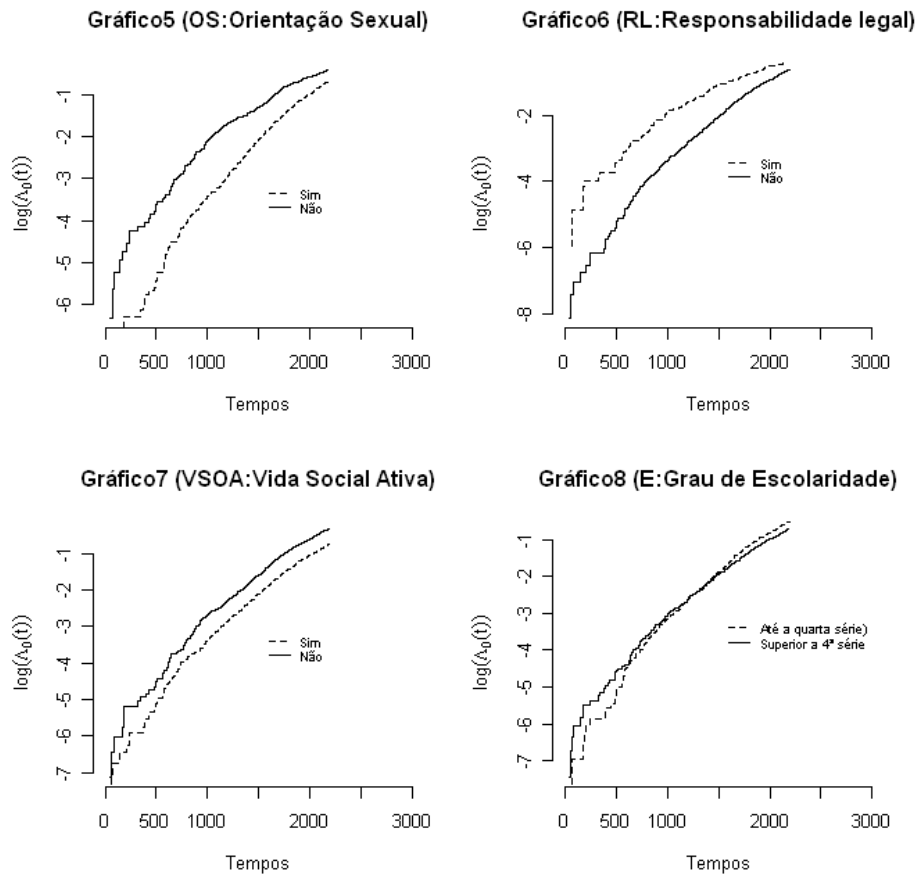


FIGURA 6 Representações gráficas do logaritmo do risco  $\text{Log}(\hat{\Lambda}_{0j}(t))$  em relação ao tempo em dias após a gravidez para as covariáveis orientação sexual (OS), responsabilidade legal do parceiro (RL), vida social ativa (VSOA) e escolaridade da gestante (E).

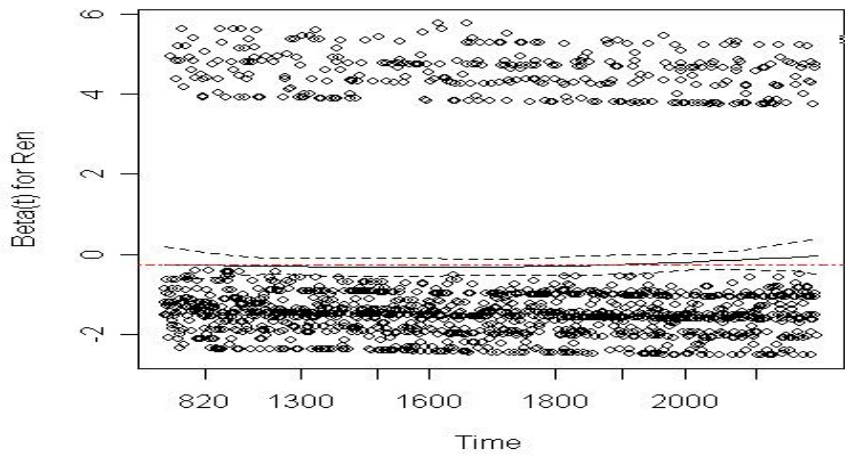


FIGURA 7 Representação gráfica da suposição de riscos proporcionais para a covariável renda (Ren).

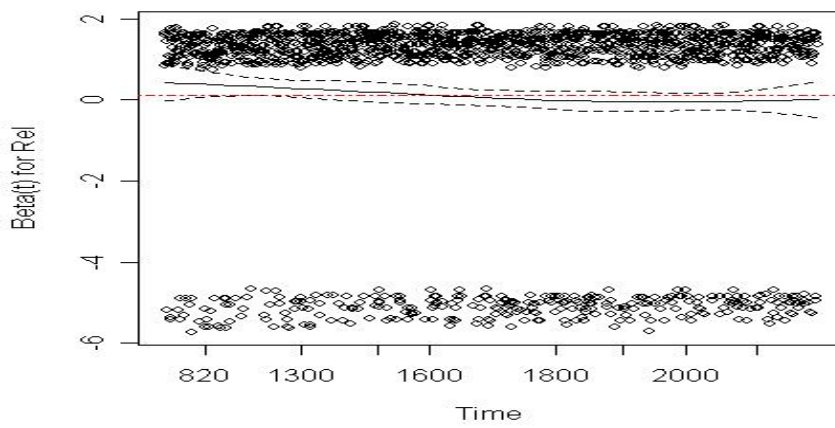


FIGURA 8 Representação gráfica da suposição de riscos proporcionais para a religião (Rel)

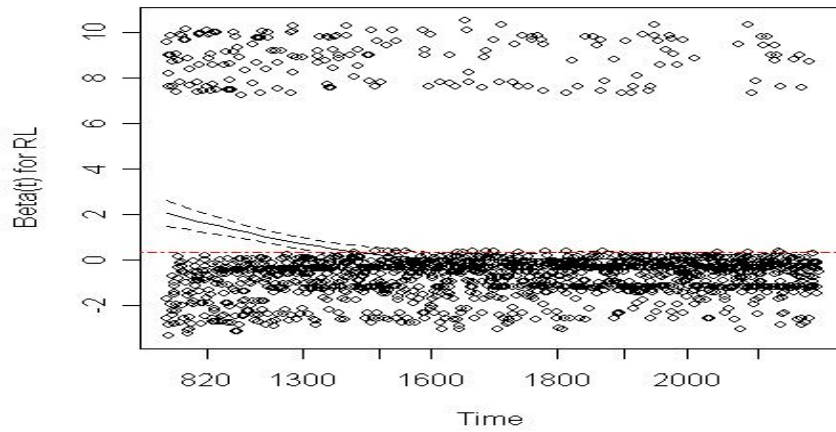


FIGURA 9 Representação gráfica da suposição de riscos proporcionais para a covariável responsabilidade legal do parceiro (RL).

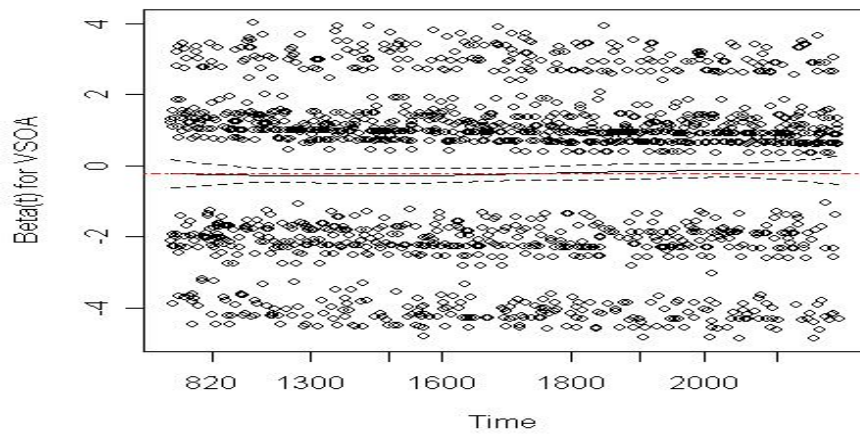


FIGURA 10 Representação gráfica da suposição de riscos proporcionais para a covariável vida social ativa (VSOA).

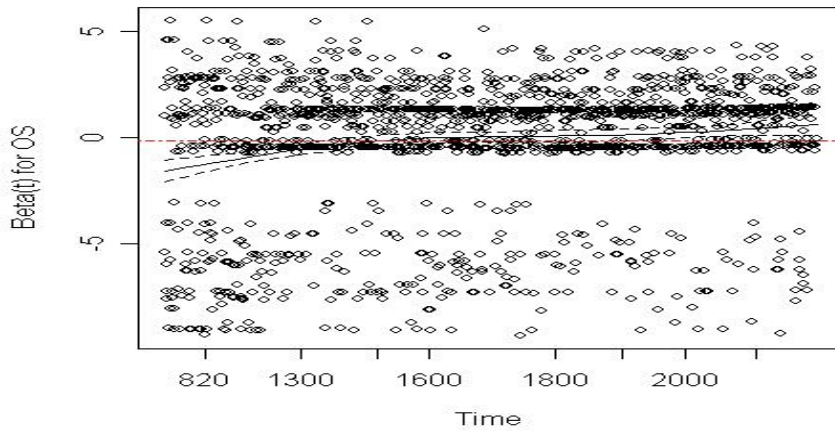


FIGURA 11 Representação gráfica da suposição de riscos proporcionais para a covariável orientação sexual (OS).

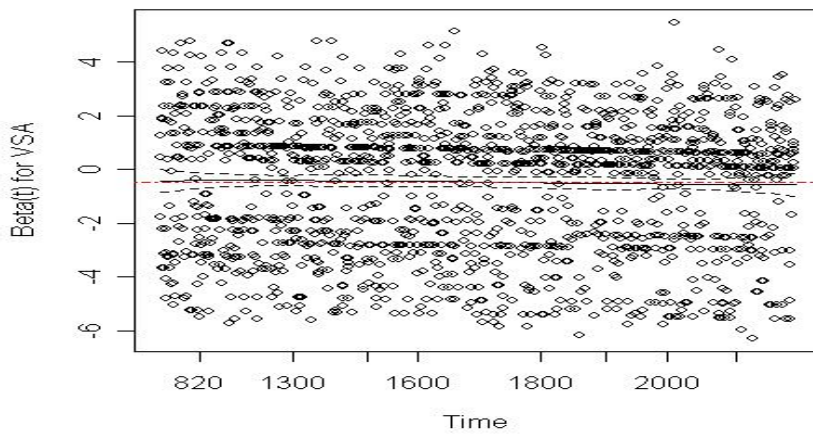


FIGURA 12 Representação gráfica da suposição de riscos proporcionais para a covariável vida sexual ativa (VSA).

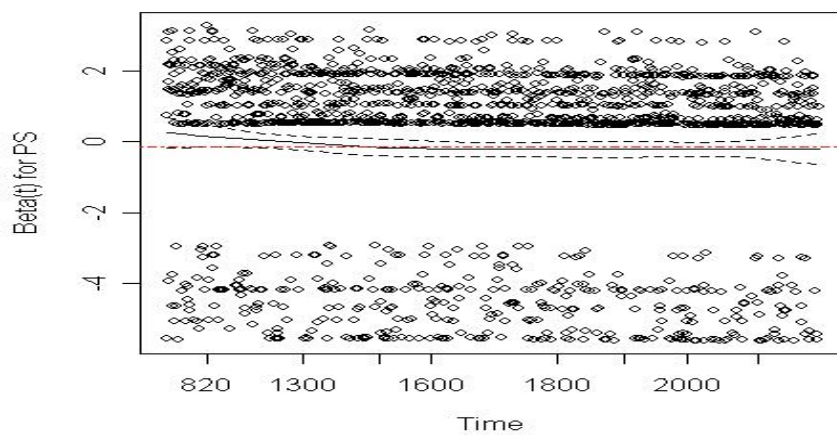


FIGURA 13 Representação gráfica da suposição de riscos proporcionais para a covariável programas sociais (PS).

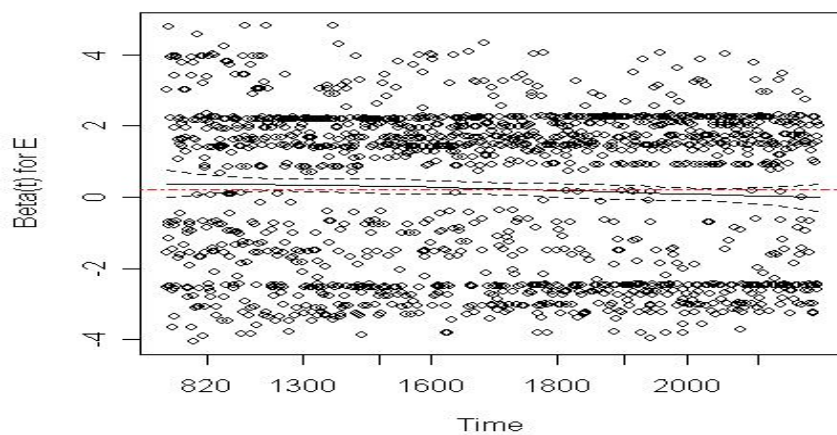


FIGURA 14 Representação gráfica da suposição de riscos proporcionais para a covariável escolaridade da gestante (E).



#### 4.8 Modelo de Cox estratificado

De acordo com a Tabela 17, quatro das covariáveis selecionadas apresentaram indícios de violação da suposição de riscos proporcionais. Uma das metodologias usadas nesse caso é fazer a estimação dos parâmetros usando o modelo de Cox estratificado. No entanto, essa metodologia não é aconselhada para situações em que mais de duas covariáveis violam o princípio da suposição de riscos proporcionais, pois os estratos podem ter tamanhos amostrais pequenos, o que pode produzir um impacto grande nas estimativas, em particular do risco de base ( $\alpha_0$ ).

TABELA 17 Estimativas dos parâmetros do testes da proporcionalidade dos riscos no modelo de Cox ajustado, valores calculados do teste Qui-quadrado e p-valor

<i>Covariável</i>	<i>Rho (<math>\rho</math>)</i>	$\chi^2_{calculado}$	<i>p-valor</i>
E	-0,0525	4,624	0,0315
OS	-0,0409	2,912	0,0879
VSA	-0,0214	0,843	0,359
OS	0,1574	45,299	<0,001
VSOA	0,0230	0,862	0,353
RL	-0,1453	34,624	<0,001
Rel	-0,0562	5,094	0,024
Ren	0,0218	0,776	0,378
GLOBAL	-	105,096	0,000

Face ao exposto e com o objetivo de reduzir o número de covariáveis que violaram a suposição de riscos proporcionais, fez-se uma comparação de vários modelos para verificar se o impacto no erro padrão das covariáveis que não violaram a suposição de riscos proporcionais foi significativo ou não.

A Tabela 18 mostra os 14 modelos avaliados. A primeira coluna é o número do modelo; a segunda, mostra os modelos que foram estratificados individualmente e a terceira mostra os modelos que foram estratificados conjuntamente. Comparando-se os resultados apresentados nas Tabelas 19 e 20, pode-se verificar que:

- a medida dos erros padrões das covariáveis PS, VSA, VSOA e Ren foram na sua maior parte menores nos modelos estratificados conjuntamente do que nos modelos estratificados individualmente;
- a covariável E apresentou um p-valor superior a 0,05 em todos os modelos;
- a covariável Rel apresentou p-valor não muito distante de 0,05;
- os modelos 5, 7, 9 e 11 tanto da estratificação individual quanto da conjunta apresentaram uma medida superior a 0,05, rejeitando-se assim a hipótese da violação da suposição de riscos proporcionais;
- o modelo 9 da estratificação conjunta apresentou um p-valor global de 0,2234, enquanto o mesmo modelo na estratificação individual apresentou um p-valor global de 0,0678 .

Depois da avaliação de modelos, optou-se pelo modelo 9 da estratificação conjunta sem interação para fazer a interpretação das estimativas dos parâmetros no modelo cox estratificado.

TABELA 18 Possíveis modelos com estratificação individual versus possíveis modelos com estratificação conjunta para obtenção do modelo ajustado de Cox estratificado para os dados do estudo da gravidez na adolescência na microrregião de Campo Belo - MG

Mod	Estratificação Individual	Estratificação Conjunta
1	Strata(E)+PS+VSA+OS+VSOA+RL+ Rel+Ren	Strata(E)+PS+VSA+OS+VSOA+RL+ Rel+Ren
2	Strata(E)+PS+VSA+strata(OS)+VSO A+RL+Rel+Ren	Strata(E+OS)+PS+VSA +VSOA+RL+Rel+Ren
3	Strata(E)+PS+VSA+OS+VSOA+Strat a(RL)+Rel+Ren	Strata(E+RL)+PS+VSA+OS+VSOA +Rel+Ren
4	Strata(E)+PS+VSA+OS+VSOA+RL+ Strata(Rel)+Ren	Strata(E+Rel)+PS+VSA+OS+VSOA+ RL +Ren
5	Strata(E)+PS+VSA+Strata(OS)+VSO A+Strata(RL)+Rel+Ren	Strata(E+OS+RL)+PS+VSA +VSOA +Rel+Ren
6	Strata(E)+PS+VSA+Strata(OS)+VSO A+RL+Strata(Rel)+Ren	Strata(E+OS+RL)+PS+VSA +VSOA +Rel+Ren
7	Strata(E)+PS+VSA+Strata(OS)+VSO A+Strata(RL)+Strata(Rel)+Ren	Strata(E+OS+RL+Rel)+PS+VSA +VSOA +Ren
8	E+PS+VSA+Strata(OS)+VSOA+RL+ Rel+Ren	E+PS+VSA+Strata(OS)+VSOA+RL+ Rel+Ren
9	E+PS+VSA+Strata(OS)+VSOA+Strat a(RL)+Rel+Ren	E+PS+VSA+Strata(OS+RL)+VSOA +Rel+Ren
10	E+PS+VSA+Strata(OS)+VSOA+RL+ Strata(Rel)+Ren	E+PS+VSA+Strata(OS+Rel)+VSOA+ RL +Ren
11	E+PS+VSA+Strata(OS)+VSOA+Strat a(RL)+Strata(Rel)+Ren	E+PS+VSA+Strata(OS+RL+Rel)+VS OA +Ren
12	E+PS+VSA+OS+VSOA+Strata(RL)+ Rel+Ren	E+PS+VSA+OS+VSOA+Strata(RL)+ Rel+Ren
13	E+PS+VSA+OS+VSOA+Strata(RL)+ Strata(Rel)+Ren	E+PS+VSA+OS+VSOA+Strata(RL+R el)+Ren
14	E+PS+VSA+OS+VSOA+RL+Strata(R el)+Ren	E+PS+VSA+OS+VSOA+RL+Strata(R el)+Ren

Comparando as Tabelas 8 e 15 e observando a Figura 15 nota-se que:

- os erros-padrão das covariáveis E, PS, VSA, VSOA, Rel e Ren não apresentaram diferenças consideráveis entre os modelos de Cox e Cox estratificado; as taxas de razão de riscos se apresentaram muito próximas para os dois modelos; a amplitude dos intervalos de confiança para as taxas de razão de risco são muito próximas para as covariáveis E e Ps e idênticas para as demais;
- nenhum dos gráficos da Figura 15 mostra séria violação à suposição de riscos proporcionais por parte das covariáveis consideradas no modelo.

TABELA 19 Resultados das estimativas das covariáveis, dos erros padrões, da razão de risco e do p- valor de vários modelos ajustados com estratificação individual para os dados do estudo da gravidez na adolescência na microrregião de Campo Belo-MG

Mod	Covariável E				Covariável PS			
	Est.	EP	RR	p-valor	Est.	EP	RR	p-valor
1	-	-	-	-	-0,129	0,0617	0,879	0,0529
2	-	-	-	-	-0,134	0,0618	0,875	0,0664
3	-	-	-	-	-0,135	0,0618	0,873	0,0861
4	-	-	-	-	-0,127	0,0618	0,881	0,0527
5	-	-	-	-	-0,142	0,0617	0,867	0,1077
6	-	-	-	-	-0,135	0,0618	0,874	0,0650
7	-	-	-	-	-0,140	0,0618	0,869	0,0991
8	0,220	0,0543	1,247	0,0529	-0,132	0,0618	0,876	0,0648
9	0,217	0,0542	1,242	0,0790	-0,138	0,0617	0,871	0,0980
10	0,222	0,0543	1,249	0,0549	-0,133	0,0618	0,876	0,0586
11	0,221	0,0543	1,248	0,0790	-0,134	0,0618	0,875	0,0815
12	0,218	0,0543	1,244	0,0596	-0,132	0,0617	0,877	0,0656
13	0,222	0,0544	1,248	0,0575	-0,127	0,0618	0,880	0,0626
14	0,220	0,0544	1,246	0,0520	-0,128	0,0617	0,880	0,0529

(Continua...)

“TABELA 19, Cont.”

Mod	Covariável VSA				Covariável OS			
	Est.	EP	Est.	EP	Est.	EP	Est.	EP
1	-0,474	0,0592	-0,474	0,0592	-0,474	0,0592	-0,474	0,0592
2	-0,485	0,0593	-0,485	0,0593	-0,485	0,0593	-0,485	0,0593
3	-0,471	0,0591	-0,471	0,0591	-0,471	0,0591	-0,471	0,0591
4	-0,475	0,0592	-0,475	0,0592	-0,475	0,0592	-0,475	0,0592
5	-0,480	0,0592	-0,480	0,0592	-0,480	0,0592	-0,480	0,0592
6	-0,485	0,0594	-0,485	0,0594	-0,485	0,0594	-0,485	0,0594
7	-0,484	0,0592	-0,484	0,0592	-0,484	0,0592	-0,484	0,0592
8	-0,483	0,0592	-0,483	0,0592	-0,483	0,0592	-0,483	0,0592
9	-0,478	0,0591	-0,478	0,0591	-0,478	0,0591	-0,478	0,0591
10	-0,481	0,0593	-0,481	0,0593	-0,481	0,0593	-0,481	0,0593
11	-0,479	0,0592	-0,479	0,0592	-0,479	0,0592	-0,479	0,0592
12	-0,469	0,0591	-0,469	0,0591	-0,469	0,0591	-0,469	0,0591
13	-0,470	0,0591	-0,470	0,0591	-0,470	0,0591	-0,470	0,0591
14	-0,476	0,0592	-0,476	0,0592	-0,476	0,0592	-0,476	0,0592
Mod	Covariável VSOA				Covariável RL			
	Est.	EP	Est.	EP	Est.	EP	Est.	EP
1	-0,212	0,0571	-0,212	0,0571	-0,212	0,0571	-0,212	0,0571
2	-0,214	0,0572	-0,214	0,0572	-0,214	0,0572	-0,214	0,0572
3	-0,211	0,571	-0,211	0,571	-0,211	0,571	-0,211	0,571
4	-0,014	0,0572	-0,014	0,0572	-0,014	0,0572	-0,014	0,0572
5	-0,213	0,0572	-0,213	0,0572	-0,213	0,0572	-0,213	0,0572
6	-0,214	0,0573	-0,214	0,0573	-0,214	0,0573	-0,214	0,0573
7	-0,210	0,0572	-0,210	0,0572	-0,210	0,0572	-0,210	0,0572
8	-0,213	0,0572	-0,213	0,0572	-0,213	0,0572	-0,213	0,0572
9	-0,209	0,0572	-0,209	0,0572	-0,209	0,0572	-0,209	0,0572
10	-0,212	0,0572	-0,212	0,0572	-0,212	0,0572	-0,212	0,0572
11	-0,207	0,0572	-0,207	0,0572	-0,207	0,0572	-0,207	0,0572
12	-0,212	0,0571	-0,212	0,0571	-0,212	0,0571	-0,212	0,0571
13	-0,211	0,0571	-0,211	0,0571	-0,211	0,0571	-0,211	0,0571
14	-0,213	0,0572	-0,213	0,0572	-0,213	0,0572	-0,213	0,0572

(Continua...)

“TABELA 19, Cont.”

Mod	Covariável Rel				Covariável Ren			
	Est.	EP	Est.	EP	Est.	EP	Est.	EP
1	0,110	0,0637	0,110	0,0637	0,110	0,0637	0,110	0,0637
2	0,105	0,0637	0,105	0,0637	0,105	0,0637	0,105	0,0637
3	0,109	0,0638	0,109	0,0638	0,109	0,0638	0,109	0,0638
4	-	-	-	-	-	-	-	-
5	0,105	0,0638	0,105	0,0638	0,105	0,0638	0,105	0,0638
6	-	-	-	-	-	-	-	-
7	-	-	-	-	-	-	-	-
8	0,105	0,0637	0,105	0,0637	0,105	0,0637	0,105	0,0637
9	0,104	0,0637	0,104	0,0637	0,104	0,0637	0,104	0,0637
10	-	-	-	-	-	-	-	-
11	-	-	-	-	-	-	-	-
12	0,105	0,0637	0,105	0,0637	0,105	0,0637	0,105	0,0637
13	-	-	-	-	-	-	-	-
14	-	-	-	-	-	-	-	-

TABELA 20 Continuação dos resultados das estimativas das covariáveis, dos erros-padrão, da razão de risco e do p-valor de vários modelos ajustados com estratificação conjunta para os dados do estudo da gravidez na adolescência em Campo Belo e Região - MG

Mod	Covariável E				Covariável PS			
	Est.	EP	RR	P-valor	Est.	EP	RR	P-valor
1	-	-	-	-	-0,129	0,0617	0,879	0,0529
2	-	-	-	-	-0,145	0,0616	0,865	0,130
3	-	-	-	-	-0,138	0,0617	0,871	0,1158
4	-	-	-	-	-0,130	0,0617	0,878	0,0505
5	-	-	-	-	-0,165	0,0613	0,848	0,4928
6	-	-	-	-	-0,151	0,0615	0,860	0,172
7	-	-	-	-	-0,166	0,0613	0,847	0,503
8	0,220	0,0543	1,247	0,0529	-0,132	0,0618	0,876	0,0648
9	0,182	0,0527	1,200	0,8018	-0,155	0,0614	0,857	0,2898
10	0,188	0,0527	1,207	0,781	-0,146	0,0615	0,864	0,185
11	0,160	0,0519	1,174	0,331	-0,157	0,0613	0,855	0,390
12	0,218	0,0543	1,244	0,0596	-0,132	0,0617	0,877	0,0656
13	0,223	0,0544	1,250	0,061	-0,126	0,0618	0,881	0,0607
14	0,220	0,0544	1,246	0,0520	-0,128	0,0617	0,880	0,0529

(Continua...)

“TABELA 20, Cont.”

Mod	Covariável VSA				Covariável OS			
	Est.	EP	Est.	EP	Est.	EP	Est.	EP
1	-0,474	0,0592	-0,474	0,0592	-0,474	0,0592	-0,474	0,0592
2	-0,507	0,0587	-0,507	0,0587	-0,507	0,0587	-0,507	0,0587
3	-0,471	0,0590	-0,471	0,0590	-0,471	0,0590	-0,471	0,0590
4	-0,477	0,0592	-0,477	0,0592	-0,477	0,0592	-0,477	0,0592
5	-0,504	0,0584	-0,504	0,0584	-0,504	0,0584	-0,504	0,0584
6	-0,519	0,0581	-0,519	0,0581	-0,519	0,0581	-0,519	0,0581
7	-0,515	0,0579	-0,515	0,0579	-0,515	0,0579	-0,515	0,0579
8	-0,483	0,0592	-0,483	0,0592	-0,483	0,0592	-0,483	0,0592
9	-0,491	0,0589	-0,491	0,0589	-0,491	0,0589	-0,491	0,0589
10	-0,504	0,0570	-0,504	0,0570	-0,504	0,0570	-0,504	0,0570
11	-0,507	0,0579	-0,507	0,0579	-0,507	0,0579	-0,507	0,0579
12	-0,469	0,0591	-0,469	0,0591	-0,469	0,0591	-0,469	0,0591
13	-0,472	0,0590	-0,472	0,0590	-0,472	0,0590	-0,472	0,0590
14	-0,476	0,0592	-0,476	0,0592	-0,476	0,0592	-0,476	0,0592
Mod	Covariável VSOA				Covariável RL			
	Est.	EP	Est.	EP	Est.	EP	Est.	EP
1	-0,212	0,0571	-0,212	0,0571	-0,212	0,0571	-0,212	0,0571
2	-0,218	0,0572	-0,218	0,0572	-0,218	0,0572	-0,218	0,0572
3	-0,211	0,0571	-0,211	0,0571	-0,211	0,0571	-0,211	0,0571
4	-0,213	0,0572	-0,213	0,0572	-0,213	0,0572	-0,213	0,0572
5	-0,216	0,0571	-0,216	0,0571	-0,216	0,0571	-0,216	0,0571
6	-0,225	0,0571	-0,225	0,0571	-0,225	0,0571	-0,225	0,0571
7	-0,222	0,0570	-0,222	0,0570	-0,222	0,0570	-0,222	0,0570
8	-0,213	0,0572	-0,213	0,0572	-0,213	0,0572	-0,213	0,0572
9	-0,214	0,0571	-0,214	0,0571	-0,214	0,0571	-0,214	0,0571
10	-0,223	0,0570	-0,223	0,0570	-0,223	0,0570	-0,223	0,0570
11	-0,225	0,0569	-0,225	0,0569	-0,225	0,0569	-0,225	0,0569
12	-0,212	0,0571	-0,212	0,0571	-0,212	0,0571	-0,212	0,0571
13	-0,212	0,0571	-0,212	0,0571	-0,212	0,0571	-0,212	0,0571
14	-0,213	0,0572	-0,213	0,0572	-0,213	0,0572	-0,213	0,0572

(Continua...)

“TABELA 20, Cont.”

Mod	Covariável Rel			Covariável Ren				
	Est.	EP	RR	Est.	EP	RR	Est.	EP
1	0,110	0,0637	1,116	0,110	0,0637	1,116	0,110	0,0637
2	0,104	0,0637	1,109	0,104	0,0637	1,109	0,104	0,0637
3	0,113	0,0637	1,119	0,113	0,0637	1,119	0,113	0,0637
4	-	-	-	-	-	-	-	-
5	0,113	0,0637	1,119	0,113	0,0637	1,119	0,113	0,0637
6	-	-	-	-	-	-	-	-
7	-	-	-	-	-	-	-	-
8	0,105	0,0637	1,110	0,105	0,0637	1,110	0,105	0,0637
9	0,111	0,0637	1,117	0,111	0,0637	1,117	0,111	0,0637
10	-	-	-	-	-	-	-	-
11	-	-	-	-	-	-	-	-
12	0,105	0,0637	1,111	0,105	0,0637	1,111	0,105	0,0637
13	-	-	-	-	-	-	-	-
14	-	-	-	-	-	-	-	-

TABELA 21 Resultado das estimativas dos parâmetros, dos erros-padrão (EP), do p-valor, da razão de risco(RR) e do intervalo de confiança da razão de risco do ajuste do modelo de Cox estratificado com estratificação conjunta (E+PS+VSA+Strata(OS+RL)+VSOA +Rel+Ren) para explicar os dados da gravidez na adolescência na microrregião de Campo Belo -MG

Covariável	Estimativa( $\beta$ )	EP	p-valor	RR( $e^\beta$ )	IC <sub>95%</sub> (RR)
E	0,182	0,0527	$5.10^{-4}$	1,200	[1,082;1,330]
PS	-0,155	0,0614	$1,2.10^2$	0,857	[0,760;0,966]
VSA	-0,491	0,0589	<0,001	0,612	[0,545;0,687]
VSOA	-0,214	0,0571	$1,8.10^4$	0,808	[0,722;0,903]
Rel	0,111	0,0637	0,03	1,117	[0,986;1,265]
Ren	-0,255	0,0626	$5.10^{-5}$	0,775	[0,685;0,876]



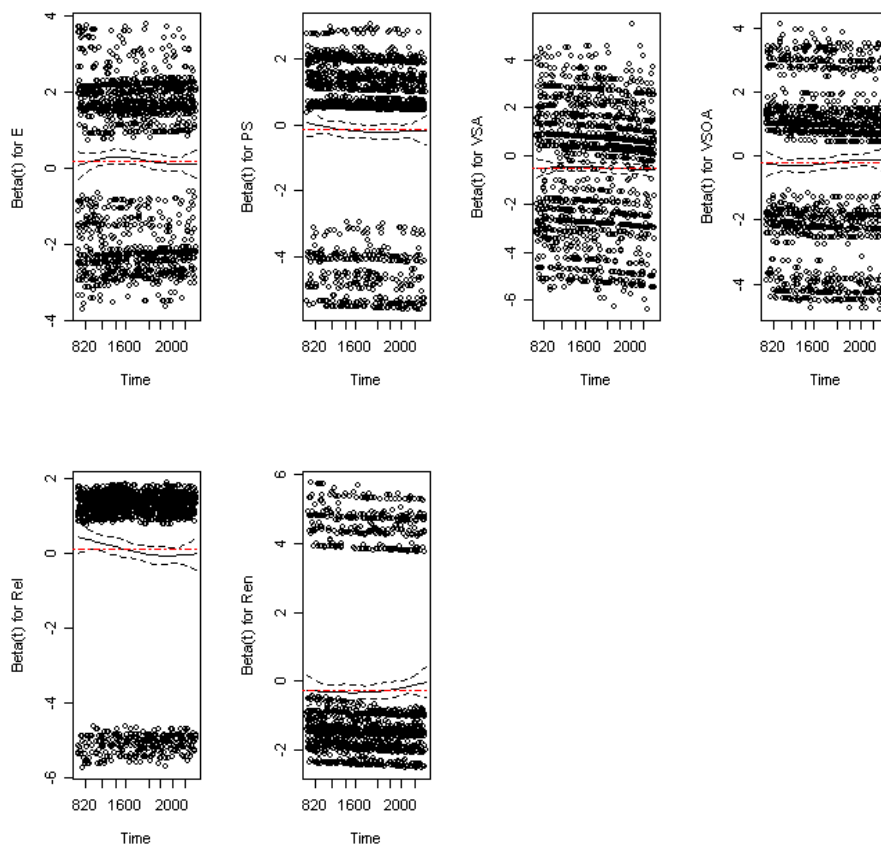


FIGURA 15 Representações gráficas dos resíduos padronizados de Schoenfeld versus os tempos para as covariáveis consideradas no modelo de Cox estratificado.

#### 4.9 Interpretação das estimativas das covariáveis no modelo Cox estratificado

As adolescentes da religião católica apresentaram 1,117 vezes maior risco de engravidar na adolescência do que as não católicas. Além disso, pode-se dizer com 95% de confiança que esse risco varia entre 0,985 e 1,265; as adolescentes filhas de pais que se separaram apresentaram 1,137 vezes maior o

risco de engravidar na adolescência do que as filhas cujos pais não se separaram. Além disso, pode-se dizer com 95% de confiança que esse risco varia entre 0,760 e 0,966.

As adolescentes que não tiveram vida sexual ativa apresentaram um risco 0,612 vezes maior de engravidar na adolescência do que aquelas que tiveram. Além disso, pode-se dizer com 95% de confiança que esse risco varia entre 0,545 e 0,687; as adolescentes que não tiveram vida social ativa apresentaram 0,808 vezes maior o risco de engravidar na adolescência do que aquelas que tiveram tal comportamento. Além disso, pode-se dizer com 95% de confiança que esse risco varia entre 0,722 e 0,903.

As adolescentes que tinham renda superior ou igual a um salário mínimo apresentaram 0,775 vezes maior o risco de engravidar na adolescência do que as com renda inferior a um salário mínimo. Além disso, pode-se dizer com 95% de confiança que esse risco varia entre 0,685 e 0,876; as adolescentes com escolaridade inferior ou igual à 4ª série primária ou ao 5º ano apresentaram 1,2 vezes maior o risco de engravidar na adolescência do que as que apresentaram escolaridade superior ao ponto de corte. Além disso, pode-se dizer com 95% de confiança que esse risco varia entre 1,082 e 1,330.

#### **4.10 Modelo de Cox com covariáveis dependentes do tempo**

Durante a coleta dos dados verificou-se que a covariável VSOA é dependente do tempo e foi possível determinar o ponto de corte de algumas observações. Com isso fez-se também uma análise usando o modelo de Cox estendido tempo-dependente, com o objetivo de encontrar o modelo que mais se ajusta aos dados.

De acordo com a Tabela 22, as covariáveis OS e RL podem estar causando violação da suposição de riscos proporcionais, pois apresentam p-valor inferiores a 5%. No entanto, os gráficos 5 e 6 da Figura 16 e os gráficos das

Figuras 17 e 18 levam a crer que as covariáveis OS e RL podem não estar causando violação da suposição de riscos proporcionais.

TABELA 22 Estimativas dos parâmetros, valores do teste Qui-quadrado e p-valor para aplicação do teste da proporcionalidade dos riscos no modelo de Cox ajustado

<i>Covariável</i>	<i>Rho (<math>\rho</math>)</i>	$\chi^2$	<i>p-valor</i>
E	-0,0303	1,566	0,211
PS	-0,0304	1,642	0,200
VSA	-0,0129	0,311	0,577
OS	0,1495	41,824	<0,0001
VSOA	0,0350	1,993	0,158
RL	-0,1636	44,996	<0,0001
Rel	-0,0435	3,102	0,082
Ren	0,0198	0,653	0,419
GLOBAL	-	116,970	0,000

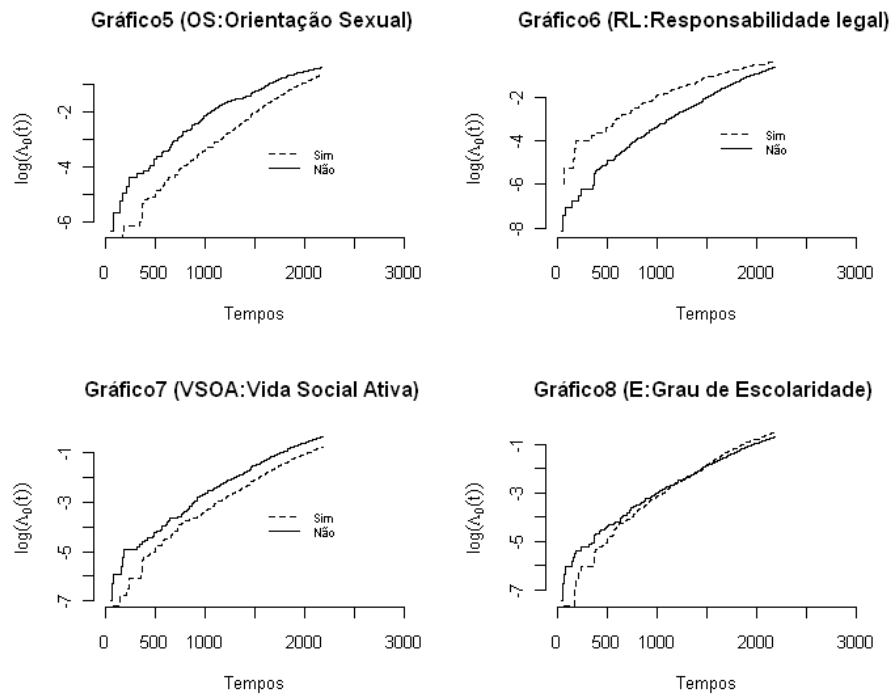


FIGURA 16 Representações gráficas dos  $\text{Log}(\hat{\Lambda}_{0j}(t))$  versus tempo para as covariáveis OS, RL, VSOA e E.

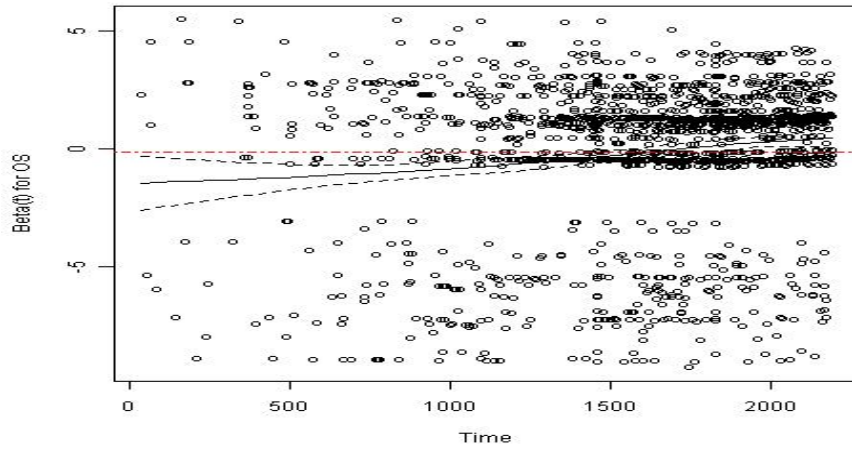


FIGURA 17 Representação gráfica da suposição de riscos proporcionais para a covariável OS.

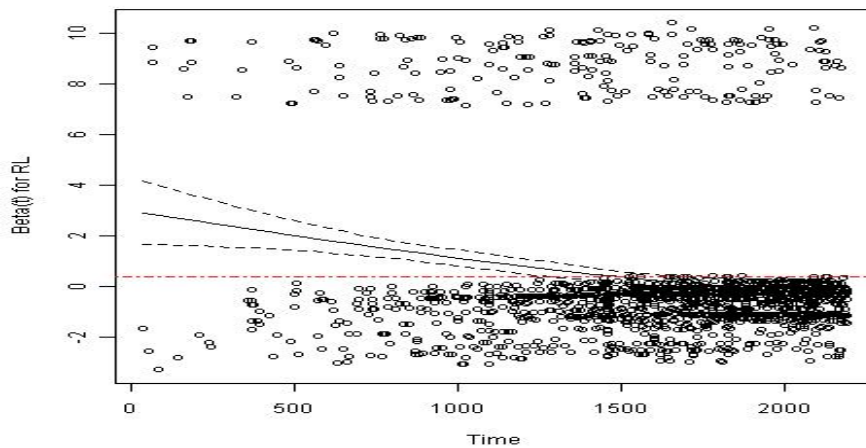


FIGURA 18 Representação gráfica da suposição de riscos proporcionais para a covariável RL.

Quanto à adequação do modelo, os gráficos dos resíduos Martingal (Figura 19) e Deviance (Figura 20) fornecem indícios favoráveis à adequação do modelo ajustado, pois mostram um comportamento aleatório em torno de zero.

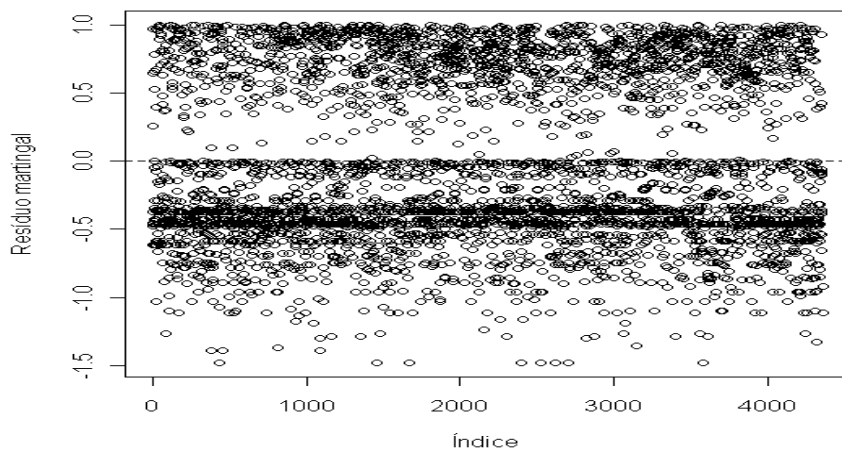


FIGURA 19 Representação gráfica dos resíduos de Martingale para o modelo covariáveis tempo-dependente.

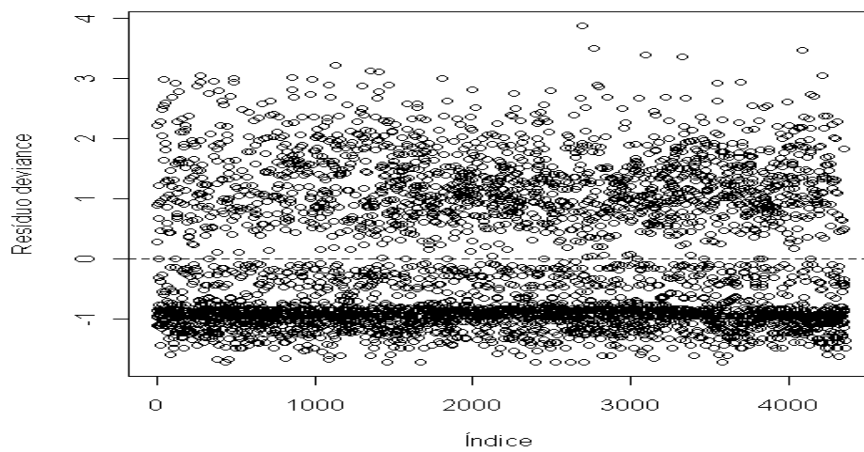


FIGURA 20 Representação gráfica resíduos de deviance para o modelo covariáveis tempo-dependente.

Os gráficos das Figuras 21 e 22 mostram que, aparentemente, não há evidências de pontos influentes no modelo ajustado.

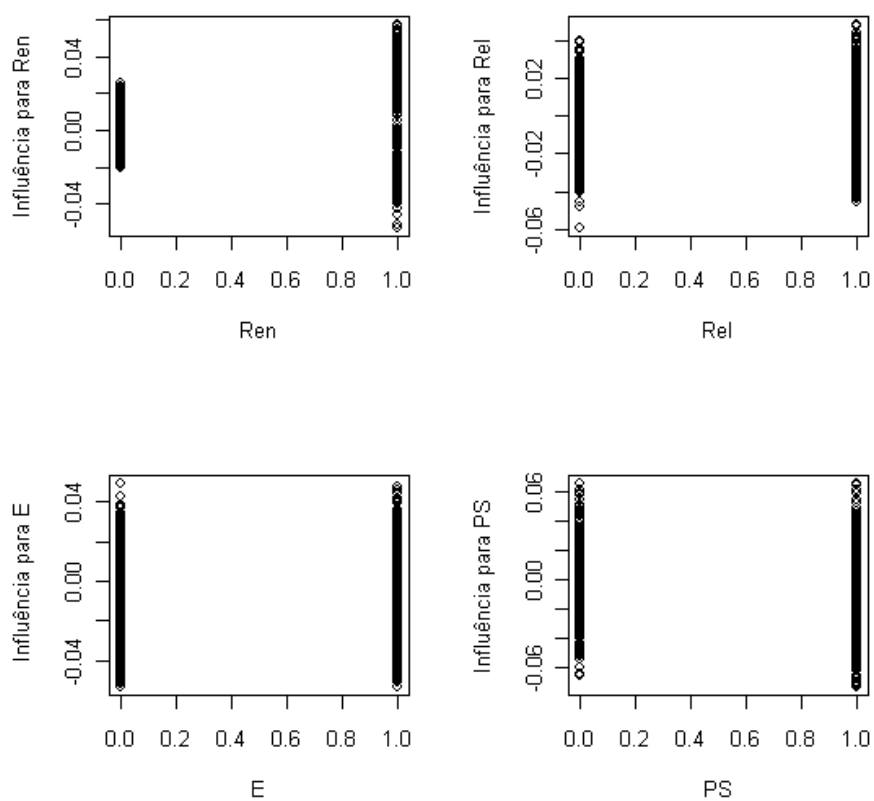


FIGURA 21 Representações gráficas dos resíduos  $dfbetas$  versus cada covariável no modelo final ajustado para os dados gravidez na adolescência usando modelo de Cox estendido covariável tempo-dependente.

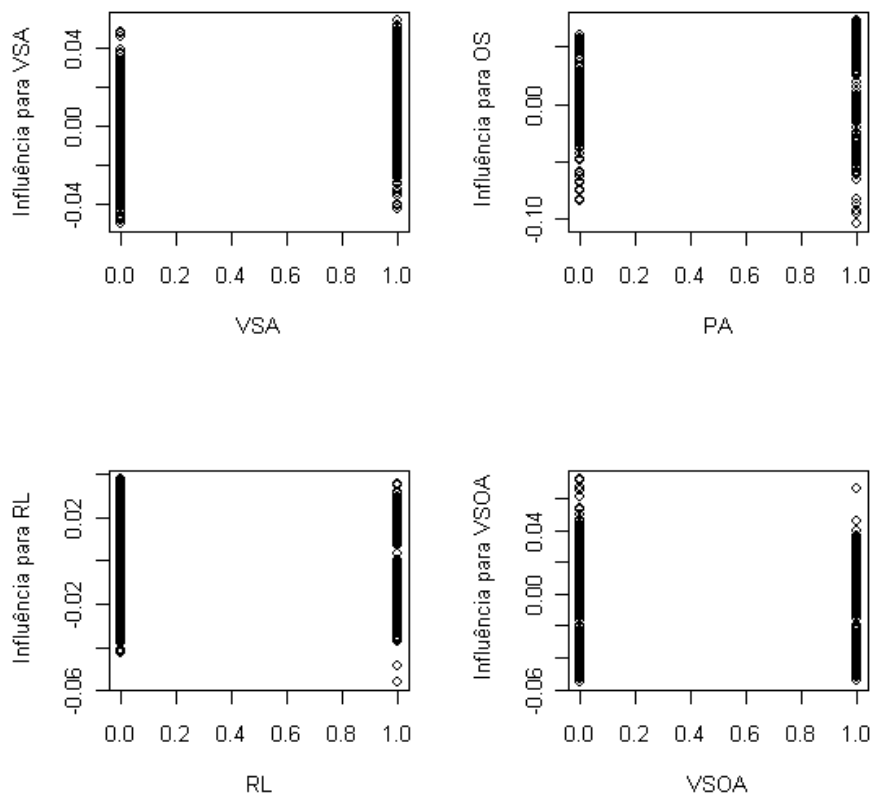


FIGURA 22 Representações gráficas dos resíduos  $dfbetas$  versus cada covariável no modelo final ajustado para os dados gravidez na adolescência usando modelo de Cox estendido covariável tempo-dependente.

De acordo com os resultados da Tabela 23, pode-se observar que: as adolescentes da religião católica apresentaram 1,080 vezes maior o risco de engravidar na adolescência do que as evangélicas e pode-se dizer com 95% de confiança que esse risco varia entre 1,017 e 1,143; as adolescentes filhas de pais não separados apresentaram 0,867 vezes maior o risco de engravidar na



adolescência do que as filhas cujos pais não se separaram e pode-se dizer com 95% de confiança que esse risco varia entre 0,806 e 0,928; as adolescentes que não tiveram vida sexual ativa apresentaram um risco 0,609 vezes maior de engravidar na adolescência do que aquelas que não tiveram, pode-se dizer com 95% de confiança que esse risco varia entre 0,552 e 0,666; as adolescentes que tiveram orientação sexual, apresentaram um risco 0,76 vezes maior de engravidar na adolescência do que aquelas que não tiveram qualquer orientação e ainda pode-se dizer com 95% de confiança que esse risco varia entre 0,804 e 0,948; as adolescentes que tiveram vida social ativa apresentaram 0,785 vezes maior o risco de engravidar na adolescência do que aquelas que não tiveram tal comportamento, pode-se dizer com 95% de confiança que esse risco varia entre 0,732 e 0,838; as adolescentes que tiveram conhecimento sobre as leis de paternidade apresentaram 1,441 vezes maior o risco de engravidar na adolescência do que as que não apresentaram ter tal conhecimento. E também, pode-se dizer com 95% de confiança que esse risco varia entre 1,362 e 1,519; as adolescentes que tinham renda superior ou igual a um salário mínimo apresentaram 0,772 vezes maior o risco de engravidar na adolescência do que as com renda inferior a um salário mínimo e pode-se dizer com 95% de confiança que esse risco varia entre 0,710 e 0,834; as adolescentes com escolaridade inferior ou igual à 4ª série primária ou ao 5º ano apresentaram 1,234 vezes maior o risco de engravidar na adolescência do que as que apresentaram escolaridade superior ao ponto de corte. Além disso, pode-se dizer com 95% de confiança que esse risco varia entre 1,180 e 1,288.

TABELA 23 Resultado das estimativas dos parâmetros, do erro padrão (EP), do p-valor, da razão de risco (RR) e do intervalo de confiança da razão de risco do ajuste do modelo de Cox com variável tempo-dependente para explicar os dados da gravidez na adolescência na microrregião de Campo Belo-MG

<i>Covariável</i>	<i>Estimativa(<math>\beta</math>)</i>	<i>EP</i>	<i>p-valor</i>	<i>RR(<math>e^\beta</math>)</i>	<i>IC<sub>95%</sub>(RR)</i>
Ren	-0,2590	0,0619	<0,0001	0,772	[0,710 ;0,834 ]
E	0,2103	0,0537	<0,0001	1,234	[1,180 ;1,288]
PS	-0,1427	0,0607	0,019	0,867	[0,806 ;0,928 ]
OS	-0,1320	0,0722	0,068	0,876	[0,804 ;0,948 ]
VSA	-0,4962	0,0567	<0,0001	0,609	[0,552;0,666 ]
RL	0,3656	0,0783	<0,0001	1,441	[ 1,362;1,519 ]
VSOA	-0,2415	0,0534	<0,0001	0,785	[ 0,732;0,838 ]
Rel	0,0768	0,0628	0,22	1,080	[ 1,017;1,143 ]

Os modelos de Cox, Cox-Estratificado, Cox-Covariável Tempo-dependente e o Paramétrico Log-Normal foram utilizados na análise dos dados da gravidez na adolescência com a premissa de selecionar as covariáveis associadas ao tempo até a gravidez. De todas as covariáveis envolvidas no estudo, os modelos paramétrico Log-Normal, Cox e Cox-covariável tempo-dependente selecionaram as mesmas covariáveis (Ren, E, PS, OS, VSA, RL, VSOA, RL) para compor o modelo ajustado. Já o modelo Cox-estratificado selecionou apenas seis das covariáveis (Ren, E, PS, VSA, RL, VSOA) selecionadas nos outros modelos. Isso mostra que as covariáveis usadas para explicar o tempo até a gravidez são as mesmas, apesar de a estrutura da modelagem ser diferente e que realmente elas são potencialmente importantes.

Observando a Tabela 24, pode-se notar que os valores dos módulos das estimativas dos coeficientes ajustados do modelo paramétrico Log-Normal e dos três modelos de Cox apontam na mesma direção, isto é, não houve diferenças acentuadas entre elas. O mesmo acontecendo com os erros-padrão.

TABELA 24 Estimativas dos coeficientes, Erro-Padrão e Exponencial das estimativas dos coeficientes para os quatros dos modelos ajustados

	P	C	CE	CVTP	P	C	CE	CVPT
Ren	0.214	-0,256	-0,255	-0,2590	0.0344	0,0626	0,0626	0,0619
E	-0.210	0,220	0,182	0,2103	0.0293	0,0544	0,0527	0,0537
PS	0.142	-0,129	-0,155	-0,1427	0.0529	0,0617	0,0614	0,0607
OS	0.209	-0,148	-	-0,1320	0.0534	0,0732	-	0,0722
VSA	0.398	-0,475	-0,491	-0,4962	0.0502	0,0592	0,0589	0,0567
RL	-0.328	0,358	-	0,3656	0.0690	0,0797	-	0,0783
VSOA	0.169	-0,213	-0,214	-0,2415	0.0491	0,0572	0,0571	0,0534
Rel	-0.124	0,110	0,111	0,0768	0.0526	0,0637	0,0637	0,0628

P: modelo paramétrico log-normal, C: modelo de Cox riscos proporcionais.

CE: modelo de Cox estratificado, CVPT: modelo de Cox covariável tempo-dependente.

Quanto à razão de risco dos três modelos de Cox, nota-se que a diferença entre elas e para cada uma das covariáveis foi muito pequena, não justificando com isso o trabalho na modelagem dos modelos de Cox-Estratificado e modelo de Cox-Variável Tempo-dependente, como mostra a Tabela 25.

TABELA 25 Tempo mediano (TM) e razão de risco (RR) das estimativas dos coeficientes dos modelos ajustados

Covariáveis	Exponenciais das estimativas dos coeficientes dos modelos				
	TM( $e^{\beta(P)}$ )	TM( $e^{-\beta(P)}$ )	RR( $e^{\beta(C)}$ )	RR( $e^{\beta(CE)}$ )	RR( $e^{\beta(CVPT)}$ )
Ren	1,239	0,807	0,774	0,775	0,772
E	0,811	1,234	1,246	1,200	1,234
PS	1,153	0,868	0,879	0,857	0,867
OS	1,232	0,811	0,863	-	0,876
VSA	1,489	0,672	0,622	0,612	0,609
RL	0,720	1,388	1,431	-	1,441
VSOA	1,184	0,845	0,808	0,808	0,785
Rel	0,883	1,132	1,116	1,117	1,080

P: modelo paramétrico log-normal, C: modelo de Cox riscos proporcionais.

CE: modelo de Cox estratificado, CVPT: modelo de Cox covariável tempo-dependente.

Os valores dos tempos medianos dos inversos das estimativas dos coeficientes do modelo paramétrico Log-Normal e a razão de risco nos modelos de Cox não apresentam diferenças significativas, como mostra a Tabela 25. Da mesma forma que os intervalos de confiança do tempo mediano dos inversos dos coeficientes das estimativas e da razão de risco dos modelos de Cox se apresentam muito similares, isto é, com limites laterais próximos para a mesma covariável nos quatros modelos e com pequenas amplitudes, como mostra a Tabela 26.

TABELA 26 Intervalos de confiança do tempo mediano das estimativas dos coeficientes do modelo paramétrico Log-Normal e da razão de risco dos modelos de Cox

Covariável	Intervalo de Confiança			
	$IC_{95\%}(e^{-\beta(P)})$	$IC_{95\%}(e^{\beta(C)})$	$IC_{95\%}(e^{\beta(CE)})$	$IC_{95\%}(e^{\beta(CVP)})$
Ren	[0,773; 0,842]	[0,685;0,875]	[0,685; 0,876]	[0,710 ;0,834 ]
E	[1,204; 1,263]	[1,120;1,386]	[1,082 ;1,330]	[1,180 ;1,288]
PS	[0,815; 0,920]	[0,779;0,992]	[0,760; 0,966]	[0,806 ;0,928 ]
OS	[0,758; 0,865]	[0,747;0,996]	-	[0,804 ;0,948 ]
VSA	[0,621; 0,722]	[0,554;0,699]	[0,545; 0,687]	[0,552;0,666 ]
RL	[1,320; 1,457]	[1,224;1,673]	-	[ 1,362;1,519 ]
VSOA	[0,795; 0,894]	[0,723;0,904]	[0,722; 0,903]	[ 0,732;0,838 ]
Rel	[1,079; 1,185]	[0,985;1,265]	[0,986;1,265]	[ 1,017;1,143 ]

P: modelo paramétrico log-normal, C: modelo de Cox riscos proporcionais.

CE: modelo de Cox estratificado, CVPT: modelo de Cox covariável tempo-dependente.

Do exposto acima, a escolha do modelo paramétrico Log-Normal ou do modelo de Cox de risco proporcionais para interpretar as estimativas das covariáveis dos dados gravidez na adolescência fica a critério do pesquisador. Como o modelo Log-Normal fornece estimativas melhores, se bem ajustado, em relação ao modelo de riscos proporcionais de Cox, sugere-se então pela escolha do modelo paramétrico Log-Normal para interpretar as estimativas das covariáveis dos dados coletados sobre a gravidez na adolescência.

## 5 CONCLUSÕES

Não houve diferença entre os modelos de regressão Log-Normal e os modelos de regressão de Cox (riscos proporcionais, estratificado e covariável dependente do tempo) para explicar os dados na gravidez na adolescência.

Usando o modelo de regressão log-Normal para explicar os dados na gravidez na adolescência na microrregião de Campo Belo pode-se dizer que as adolescentes católicas apresentam 13% a mais de possibilidade de engravidar na adolescência do que as evangélicas; as adolescentes que tiveram algum tipo de orientação sexual apresentaram 18,8% a menos de possibilidade de engravidar na adolescência do que as que não tiveram nenhum tipo de orientação; as adolescentes que não tiveram uma vida sexual ativa apresentaram 33% a menos de possibilidade de não engravidar em relação as que tiveram; as adolescentes que tiveram conhecimento sobre a lei de paternidade apresentaram 38,8% a mais de possibilidade de engravidar na adolescência do que as que desconheciam tal lei. As adolescentes que não tiveram uma vida social ativa apresentaram 15,54% a menos de possibilidade de engravidar do que as que tiveram; as adolescentes que apresentaram escolaridade até 4ª série apresentaram 23,3% a menos de possibilidade de engravidar na adolescência em relação às com escolaridade inferior; as adolescentes com renda inferior ou igual a um salário mínimo apresentaram 19,3% a mais de possibilidade de engravidar em relação àquelas com salário superior ao mínimo e que as adolescentes que tiveram os pais separados apresentaram 13,2% a mais a possibilidade de engravidar do que as que não tiveram.

## 6 REFERÊNCIAS BIBLIOGRÁFICAS

AQUINO, E. M. L.; HEILBORN, M. L.; KNAUTH, D.; BOZON, M.; ALMEIDA, M. da C.; ARAÚJO, J.; MENEZES, G. Adolescência e reprodução no Brasil: a heterogeneidade dos perfis sociais. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 19, n. 2, p. 377-388, jan. 2003. Suplemento.

BARALDI, A. C. P.; DAUD, Z. P.; ALMEIDA, A. M. de; GOMES, F. A.; NAKANO, A. M. S. Adolescent pregnancy: a comparative study between mothers who use public and private health systems. **Revista Latino-Americana de Enfermagem**, Ribeirão Preto, v. 15, p. 799-805, set./out. 2007. Edição especial.

BELO, M. A. V.; SILVA, J. L. P. e. Conhecimento, atitude e prática sobre métodos anticoncepcionais entre adolescentes gestantes. **Revista de Saúde Pública**, São Paulo, v. 38, n. 4, p. 479-487, ago. 2004.

BERLOFI, L. M.; ALKMIN, E. L. C.; BARBIERI, M.; GUAZZELLI, C. A. F.; ARAÚJO, F. F. de. Prevenção da reincidência de gravidez em adolescentes: efeitos de um Programa de Planejamento Familiar. **Acta Paulista de Enfermagem**, São Paulo, v. 19, n. 2, p. 196-200, abr./jun. 2006.

BORGES, A. L. V.; SCHOR, N. Trajetórias afetivo-amorosas e perfil reprodutivo de mulheres adolescentes residentes no Município de São Paulo. **Revista Brasileira de Saúde Materno Infantil**, Recife, v. 5, n. 2, p. 163-170, abr./jun. 2005.

BRASIL. Ministério da Saúde. **DATASUS**. Disponível em: <<http://w3.datasus.gov.br/datasus/index.php>>. Acesso em: 10 jan. 2009.

BRASIL. Ministério da Saúde. **Saúde Brasil 2006**: uma análise da desigualdade em saúde. Brasília, 2006. 150 p.

CABRAL, C. S. Contraceção e gravidez na adolescência na perspectiva de jovens pais de uma comunidade favelada do Rio de Janeiro. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 19, n. 2, p. 283-292, jan. 2003. Suplemento.

CAPUTO, V. G.; BORDIN, I. A. Gravidez na adolescência e uso freqüente de álcool e drogas no contexto familiar. **Revista de Saúde Pública**, São Paulo, v. 42, n. 3, p. 402-410, jun. 2008.

CARVALHO, M. A. B. D.; SOUZA, D. V.; OLIVEIRA, D. S.; PIRES, L. P. M. Gravidez na adolescência: aspectos socioeconômicos. **Revista de Ginecologia e Obstetria**, São Paulo, v. 8, n. 1, p. 7-14, jan. 1999.

CARVALHO, M. S.; ANDREOZZI, V. L.; CODEÇO, C. T.; BARBOSA, M. T. S.; SHIMAKURA, S. E. **Análise de sobrevivência**: teoria e aplicações em saúde. Rio de Janeiro: FIOCRUZ, 2005. 296 p.

CHALEM, E.; MITSUHIRO, S. S.; FERRI, C. P.; BARROS, M. C. M.; GUINSBURG, R.; LARANJEIRA, R. Gravidez na adolescência: perfil sócio-demográfico e comportamental de uma população da periferia de São Paulo, Brasil. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 23, n. 1, p. 177-186, jan. 2007.

COLLET, D. **Modelling survival data in medical research**. New York: Chapman and Hall, 1994. 331 p.

COLOSIMO, E. A.; GIOLO, S. R. **Análise de sobrevivência aplicada**. São Paulo: E.Blücher, 2006. 370 p.

COX, D. R. Partial likelihood. **Biometrika**, London, v. 62, n. 2, p. 269-276, 1975.

COX, D. R.; OAKES, D. **Analysis of survival data**. London: Chapman and Hall, 1984. 201 p.

COX, D. R.; SNELL, E. J. A general definition of residuals. **Journal of the Royal Statistical Society B**, London, v. 30, n. 1, p. 248-275, 1968.

ELFENBEIN, D. S.; FELICE, M. E. Adolescent pregnancy. **Pediatric Clinics of North America**, Philadelphia, v. 50, n. 4, p. 781-800, Aug. 2003.

ESTEVEZ, J. R.; MENANDRO, P. R. M. Trajetórias de vida: repercussões da maternidade adolescente na biografia de mulheres que viveram tal experiência. **Estudos de Psicologia**, Natal, v. 10, n. 3, p. 363-370, set./dez. 2005.

GUIMARÃES, E. B. **Gravidez na adolescência fatores de risco**. São Paulo: Atheneu, 2001. 298 p.

KALBFLEISCH, J. D.; PRENTICE, R. L. **The statistical analysis of failure time data**. 2. ed. Hoboken: J.Wiley, 1980. 439 p.

KAPLAN, E. L.; MÉIER, P. Nonparametric estimation from incomplete observations. **Journal of the American Statistical Association**, New York, v. 53, n. 282, p. 457-481, June 1958.

KLEINBAUM, D. G.; KLEIN, M. **Survival analysis: a self-learning text**. New York: Springer, 1997. 590 p.

LAWLESS, J. F. **Statistical models and methods for lifetime data**. New York: J.Wiley, 1982. 580 p.



MENEZES, I. H. C. F.; DOMINGUES, M. H. M. da S. Principais mudanças corporais percebidas por gestantes adolescentes assistidas em serviços públicos de saúde de Goiânia. **Revista de Nutrição**, Campinas, v. 17, n. 2, p. 185-194, abr./jun. 2004.

MICHELAZZO, D.; YAZLLE, M. E. H. D.; MENDES, M. C.; PATTA, M. C.; ROCHA, J. S. Y.; MOURA, M. D. de. Indicadores sociais de grávidas adolescentes: estudo caso-controle. **Revista Brasileira de Ginecologia e Obstetria**, Rio de Janeiro, v. 26, n. 8, p. 633-639, set. 2004.

MONTEIRO, C. F. de S.; COSTA, N. S. S.; NASCIMENTO, P. S. V. A violência intra-familiar contra adolescentes grávidas. **Revista Brasileira de Enfermagem**, Brasília, v. 60, n. 4, p. 373-376, jul./ago. 2007.

PARAGUASSU, A. L. C. B.; COSTA, M. C. O.; NASCIMENTO SOBRINHO, C. L.; PATEL, B. N.; FREITAS, J. T. de; ARAÚJO, F. P. O. de. Situação sociodemográfica e de saúde reprodutiva pré e pós-gestacional de adolescentes, Feira de Santana, Bahia, Brasil. **Ciência & Saúde Coletiva**, Rio de Janeiro, v. 10, n. 2, p. 373-380, abr./jun. 2005.

R DEVELOPMENT CORE TEAM. **The R manuals**: current. Version 1.9.1. Genebra, 2008. Disponível em: <<http://www.r-project.org>>. Acesso em: 10 mar. 2009.

ROCHA, C. S. **Modelos com fragilidade em análise de sobrevivência**. 1995. 184 f. Tese (Doutorado em Probabilidade e Estatística) - Faculdade de Ciências, Universidade de Lisboa, Lisboa.

RODRIGUES, A. P.; SOUZA, M. B.; BRASIL, R. M. C.; CARAKUSHANSKY, G. Gravidez na adolescência. **Femina**, Rio de Janeiro, v. 21, n. 1, p. 199-223, jan. 1993.

SABROZA, A. R.; LEAL, M. do C.; GAMA, S. G. N. da; COSTA, J. V. da. Perfil sócio-demográfico e psicossocial de puérperas adolescentes do Município do Rio de Janeiro, Brasil - 1999-2001. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 20, n. 1, p. 120-122, jan. 2004.

SANTOS, G. H. N. dos; MARTINS, M. da G.; SOUSA, M. da S. Gravidez na adolescência e fatores associados ao baixo peso ao nascer. **Revista Brasileira de Ginecologia e Obstetrícia**, Rio de Janeiro, v. 30, n. 5, p. 224-231, maio 2008.

SILVA, N. C. B. da; BONFIM, T.; CARDOZO, N. P.; FRANCO, M. A. P.; MARQUES, S. L. Proposta de instrumento para avaliar conhecimento de jovens sobre métodos contraceptivos. **Paidéia**, Ribeirão Preto, v. 17, n. 38, p. 365-374, set./dez. 2007.

SIMÕES, V. M. F.; SILVA, A. A. M. da; BETTIOL, H.; LAMY FILHO, F.; TONIAL, S. R.; MOCHEL, E. G. Características da gravidez na adolescência em São Luís, Maranhão. **Revista de Saúde Pública**, São Paulo, v. 37, n. 5, p. 559-565, out. 2003.

SOUZA, A. I.; AQUINO, M. M. A.; CECATTI, J. G.; SILVA, J. L. P. e. Epidemiologia do abortamento na adolescência. **Revista Brasileira de Ginecologia e Obstetrícia**, Rio de Janeiro, v. 21, n. 3, p. 161-165, abr. 1999.

STRASBURGER, V. C. **Os adolescentes e a mídia**: impacto psicológico. Porto Alegre: Artes Médicas Sul, 1999. 130 p.

THERNEAU, T. M.; GRAMBSCH, P. M. **Modeling survival data**: extending the Cox model. New York: Springer-Verlag, 2000. 332 p.

VALENÇA, D. M. **O modelo de regressão gama generalizada para discriminar entre modelos paramétricos de tempo de vida**. 1994. 134 f. Dissertação (Mestrado em Estatística) - Universidade Estadual de Campinas, Campinas.

WORLD HEALTH ORGANIZATION. **International classification of diseases**. 9. ed. Geneva, 1975. 300 p.

XIMENES NETO, F. R. G.; DIAS, M. do S. de A.; ROCHA, J.; CUNHA, I. C. K. O. Gravidez na adolescência: motivos e percepções de adolescentes. **Revista Brasileira de Enfermagem**, Brasília, v. 60, n. 3, p. 279-285, maio/jun. 2007.

YAZLLE, M. E. H. D. Gravidez na adolescência. **Revista Brasileira de Ginecologia e Obstetria**, Rio de Janeiro, v. 28, n. 8, p. 443-445, ago. 2006.

YOUNG, E. W.; JENSEN, L. C.; OLSEN, J. A.; CUNDICK, C. The effects of family structure on the sexual behavior of adolescents. **Adolescence**, San Diego, v. 26, n. 104, p. 977-986, 1991.

## ANEXO

ANEXO A	Página
TABELA 1A Descrição das Covariáveis usadas nos estudos da gravidez na adolescência na migroregião de Campo Belo - MG.....	124

## ANEXO A

TABELA 1A Descrição das Covariáveis usadas nos estudos da gravidez na adolescência na migroregião de Campo Belo - MG

Covariável	Descrição
temp	Tempo
cens	Censura: 0 se engravidou na adolescência; 1 caso contrário
Ren	Renda: 0 se inferior ou igual a um salário mínimo; 1 caso contrário
Rel	Religião: 0 se for católica; 1 evangélica
M	Mídia: 0 assiste novela; 1 outros programas
Cor	Cor: 0 afrodecentende; 1 branca
EP	Escolaridade do pai; 0 até o 5º Ano; 1 acima do 5º Ano
EM	Escolaridade da mãe; 0 até o 5º Ano; 1 acima do 5º Ano
E	Escolaridade da adolescente; 0 até o 5º Ano; 1 acima do 5º Ano
MP	Morou com os pais; 0 sim; 1 não
PS	Pais separados: 0 sim; 1 não
OS	Orientação sexual; 0 sim; 1 não
VSA	Vida sexual ativa: 0 sim; 1 não
RL	Responsabilidade legal do parceiro: 0 sim; 1 não
PA	Programas assistências a gestante: 0 sim; 1 não
MC	Método contraceptivo: 0 sim; 1 não
VSOA	Vida social ativa: 0 sim; 1 não
D	Dialogo com os pais sobre sexo, etc: 0 sim; 1 não
Dr	Uso de drogas na adolescência: 0 sim; 1 não
OM	Obtenção dos anticoncepcionais: 0 sim; 1 não