



**UNIVERSIDADE DE ÉVORA**

**ESCOLA DE CIÊNCIAS E TECNOLOGIA**

**Mestrado em Modelação Estatística e Análise de Dados**  
**Especialização em Modelação Estatística e Análise de Dados**

**Dissertação**

**Uma Abordagem Estatística da Fecundidade em Portugal**  
**(1995-2009)**

Lídia Patrícia Santos Amaral Tomé

**Orientador:**

Professora Doutora Maria Filomena Ferreira Mendes

**Co-Orientador:**

Professor Doutor Paulo Jesus Infante Santos

*“Esta Dissertação não inclui as críticas e sugestões feitas pelo Júri”*

16-08-2011

**Mestrado em Modelação Estatística e Análise de Dados**  
**Especialização em Modelação Estatística e Análise de Dados**

**Dissertação**

**Uma Abordagem Estatística da Fecundidade em Portugal**  
**(1995-2009)**

Lídia Patrícia Santos Amaral Tomé

**Orientador:**

Professora Doutora Maria Filomena Ferreira Mendes

**Co-Orientador:**

Professor Doutor Paulo Jesus Infante Santos

*“Esta Dissertação não inclui as críticas e sugestões feitas pelo Júri”*

## Resumo

O acentuado declínio e adiamento da fecundidade surge associado muitas vezes a questões como o prolongamento da formação e do ensino e a forte participação feminina no mercado de trabalho. Estas questões demográficas são hoje preocupação de cientistas e políticos que procuram evitar o acentuado envelhecimento da população, nomeadamente, em países do Sul da Europa como Portugal.

O objectivo deste trabalho é assim o de dar resposta a questões relacionadas com o adiamento e a diminuição dos níveis de fecundidade em Portugal no período entre 1995 e 2009. Numa perspectiva estatística procurou-se encontrar factores influenciadores do espaçamento entre o casamento e o nascimento do primeiro filho; entre o nascimento do primeiro e segundo filho; entre o nascimento do segundo e terceiro filho, e procurou-se também compreender que factores são responsáveis pela diminuição do número de filhos, utilizando metodologias estatísticas como análise de sobrevivência e Modelos Lineares Generalizados.

Observou-se que são factores como o nível de instrução da mãe, a situação profissional ou a condição perante o trabalho da mãe que se encontram marcadamente presentes na compreensão dos diferentes comportamentos de fecundidade analisados ao longo deste trabalho.

## A statistical study of fertility in Portugal (1995-2009)

### Abstract

The sharp decline of fertility and the postponement are often associated with issues such as the educational level or the strong female participation in the labor market. These demographic questions are nowadays concerns of scientists and politicians who seek to avoid the population aging, with particular attention in Southern European countries like Portugal.

The fundamental objective of this thesis is understand some of the issues related to delay and decreased of fertility levels in Portugal for the period of 1995-2009. From a statistical perspective sought to find factors influencing the spacing between marriage and first birth, between the birth of the first and second child, between birth of the second and third child, and attempts were made to understand which factors are responsible for the decrease on fertility, using statistical methods such as survival analysis and general linear models.

It was observed that are factors such as mother's education level, employment status or labor status of the mother that are present in the understanding of fertility behavior analyzed throughout this thesis.

## Índice

1.	Contextualização .....	6
1.1.	Introdução.....	6
1.2.	Breve revisão da literatura.....	8
1.2.1.	Impacto da Educação na fecundidade .....	12
1.2.2.	Participação no Mercado de Trabalho .....	14
2.	Análise da fecundidade.....	17
2.1.	Dados .....	17
2.2.	Análise descritiva .....	17
2.3.	Indicadores demográficos.....	24
2.3.1.	Metodologia .....	25
2.3.2.	Resultados .....	26
3.	Análise de Sobrevivência .....	30
3.1.	Metodologia.....	30
3.1.1.	Censura & Truncatura .....	32
3.1.2.	Modelo de Cox .....	33
3.1.3.	Estimador de Kaplan-Meier.....	36
3.1.4.	Modelo Aditivo de Aalen.....	37
3.2.	Resultados.....	39
3.2.1.	Aplicação do Modelo de Cox e Modelo de Cox estratificado.....	39
3.2.1.1.	Espaçamento entre o casamento e o nascimento do primeiro filho.....	39
3.2.1.2.	Espaçamento entre o nascimento do primeiro e o segundo filho.....	57
3.2.1.3.	Espaçamento entre o nascimento do segundo e o terceiro filho .....	75
3.2.2.	Aplicação do modelo aditivo de Aalen .....	91
4.	Modelos Lineares Generalizados .....	92
4.1.	Metodologia.....	92
4.1.1.	Modelo Poisson, Quasi-Poisson e Binomial-Negativo.....	94
4.1.2.	Pressupostos associados aos Modelos lineares generalizados .....	96
4.2.	Resultados.....	97
4.2.1.	Modelo de Poisson e Modelo Quasi-Poisson .....	98
4.2.2.	Modelo Binomial-Negativo .....	100
5.	Discussão dos resultados.....	107
6.	Bibliografia.....	111
	Anexo A .....	115
	Anexo B.....	127

## **1. Contextualização**

### **1.1. Introdução**

A situação da fecundidade em Portugal é caracterizada pelo declínio verificado nas últimas décadas sendo as perspectivas futuras preocupantes. Face a esta situação tem-se verificado o aumento da preocupação dos governos e o interesse dos cientistas relativamente a esta questão. No entanto, é notória a ausência de análises puramente estatísticas, aprofundadas, sobre o assunto que permitam compreender melhor os factores que determinaram a evolução passada e as perspectivas para o futuro. Desta forma, a motivação para elaboração desta dissertação, no âmbito do Mestrado em Modelação Estatística e Análise de dados, é o de dar resposta à necessidade de uma análise estatística da fecundidade em Portugal, mais aprofundada, conciliando o que foi aprendido neste Mestrado com a aprendizagem passada resultado da formação anterior.

Para a melhor compreensão do fenómeno, centrámos o nosso estudo no comportamento da fecundidade observada em Portugal, no período compreendido entre 1995 e 2009, partindo da análise dos nascimentos registados. Desta forma definimos 3 objectivos fundamentais: (1) compreender a evolução da fecundidade em Portugal no período considerado, (2) avaliar o impacto no declínio da fecundidade da redução do número de filhos dos casais e do aumento da idade média em que as mulheres têm os seus filhos, (3) avaliar quais os factores que podem explicar, por um lado, a redução no número de filhos e, por outro, o adiamento/espacamento do seu nascimento.

Sendo o tema principal desta dissertação a caracterização da actual situação de fecundidade em Portugal, neste primeiro capítulo será realizada a contextualização da situação demográfica, nomeadamente no que diz respeito à fecundidade e às metodologias aplicadas ao estudo desta covariável micro demográfica. Proceder-se-á também a uma explicação dos dados utilizados para este trabalho, acompanhada da justificação para a sua utilização.

No segundo capítulo será realizada uma análise descritiva das covariáveis em análise, de forma a compreendermos a evolução do comportamento de fecundidade em Portugal. Este capítulo será complementado também pela explicação metodológica dos

indicadores demográficos tradicionalmente utilizados para análise da situação da fecundidade, seguido de uma breve análise dos resultados obtidos.

No terceiro capítulo desenvolver-se-á uma abordagem estatística mais aprofundada com a utilização da *Análise de Sobrevivência* para analisar o espaçamento decorrido entre o casamento e o primeiro filho, entre o primeiro e o segundo filho e, por fim, entre o segundo e o terceiro filho. Será dada uma especial atenção ao *Modelo de Cox*, *Modelo de Cox estratificado* e ao *Modelo Aditivo de Aalen*.

No quarto capítulo desenvolver-se-á uma outra abordagem estatística, com utilização de *Modelos Lineares Generalizados*, usados com o objectivo de compreender o que influencia o aumento ou a diminuição do número de filhos.

Terminaremos este trabalho com as conclusões finais e propostas para futura pesquisa no âmbito da estatística bem como da demografia.

## 1.2. Breve revisão da literatura

A demografia é a ciência que estuda a população humana, a sua dimensão, distribuição, composição, bem como os factores que determinam as alterações naquelas componentes. Assim, e para compreender e analisar a problemática da demografia, é necessário o acesso a metodologias adequadas, nomeadamente metodologias estatísticas, que impulsionaram durante os anos 70 a aproximação das duas ciências (Siegel et al. 2004). É assim importante fazer uma breve contextualização da análise demográfica recente, bem como da evolução das questões relacionadas com o tema base deste trabalho, a fecundidade.

Desde meados do século XIX até à segunda metade do século XX, o modelo pós-malthusiano, caracterizado por famílias de maiores dimensões que as actuais, era o modelo prevalente em todos os países europeus, incluindo Portugal. A alteração de um modelo com 3 e mais filhos para um modelo que varia entre 1 e 2 filhos, é o resultado das constantes mudanças na sociedade, na economia, na política e mesmo no desenvolvimento cultural dos países. sociedade, na economia, na política e mesmo no desenvolvimento cultural dos países. A estas alterações na sociedade estão associados, em primeiro lugar, factores como a melhoria nas condições de vida, o avanço no saneamento público e privado, que entre outros factores contribuíram para a diminuição dos níveis de mortalidade, nomeadamente da mortalidade infantil.

O período pós-malthusiano, referido na literatura como a primeira transição demográfica, apresenta as características acima referidas e permitiu um crescimento exponencial das populações. Na sua base mais essencial esta transição demográfica é caracterizada por um elevado número de nascimentos, bem como uma mortalidade muito elevada, passando posteriormente para um acentuado declínio dos nascimentos e valores da mortalidade que são mais baixos que os seus antecedentes. Foi uma transição caracterizada por medidas altruístas e atitudes incentivadoras da união familiar, que depressa se transformaram, resultado da industrialização, no urbanismo desenfreado e na secularização, que se verificava um pouco por toda a Europa.

A diminuição das famílias, levou também a que “valor económico” das crianças enquanto mão-de-obra deixasse de existir, da mesma forma que o uso de métodos contraceptivos se tornou mais abrangente, o que directa ou indirectamente levou ao declínio da idade média do casamento assim como fez diminuir também o número de

peças não casadas (Matsuo 2003, Sobotka 2003). Esta evolução de comportamento descrito anteriormente resultou no que alguns autores (Van de Kaa 2002, Lesthaeghe e Neels 2002) defendem ser a segunda transição demográfica. Nos finais dos anos 90 do século passado, diferentes teóricos franceses observaram uma mudança notável na sua população, o número de crianças por família diminuíra significativamente como resultado directo dos esforços deliberados para reduzir a fecundidade dentro e fora do casamento.

Depressa se compreendeu que a limitação voluntária da fecundidade no casamento deveria ser considerada uma revolução demográfica, alterando os comportamentos da maioria da população e modificando simultaneamente a dinâmica da mesma. Este entendimento diferente do declínio da fecundidade era um ajustamento necessário pelo anteriormente verificado declínio da mortalidade. (Van de Kaa 2002). Se na primeira transição demográfica a principal característica foi a diminuição acentuada da mortalidade, já no caso da segunda transição demográfica a principal característica é o declínio da fecundidade e da nupcialidade. De acordo com Matsuo (2003), a primeira transição demográfica acontece aquando da estabilização dos elevados valores da mortalidade, e a segunda transição aquando da diminuição da fecundidade levando simultaneamente a níveis de reposição da população inferiores ao necessário.

Por outro lado, e já nos anos 60, verificou-se uma drástica modificação no comportamento da formação familiar e do casamento, nomeadamente, no Norte da Europa. A idade ao primeiro casamento aumentou depois de um declínio sem precedentes que durou até 1960, ao mesmo tempo que aumentou o número de divórcios assim como a coabitação. E foi também na década de 60 que se começou a verificar o adiamento da fecundidade, como consequência destes novos modos de vida. Já na década de 70, essencialmente nos países da Europa Ocidental, registou-se um adiamento que apesar de diferenciado apresentava algumas semelhanças entre os diversos países, levando a que mais tarde na década de 90, os valores nacionais do índice sintético de fecundidade<sup>1</sup> (*ISF*) reflectissem o adiamento da idade de reprodução da mulher que se encontra perto dos 30 anos de idade na maioria dos países (Lesthaeghe e Surkyn 1998).

Já nos finais da década de 80, nos países do Sul da Europa, inicia-se a segunda transição demográfica. Itália, Portugal e Espanha, iniciando o adiamento do casamento e da

---

<sup>1</sup> O índice sintético de fecundidade, indica o número de filhos que em média uma mãe deixa na população se as condições do momento se mantiverem estáveis.

fecundidade, contudo, com diferentes características em ambos os factores; isto é, a coabitação e o nascimento de filhos fora do casamento aconteceram de forma diferente, falhando ou progredindo lentamente no caso da Itália, e no caso Ibérico (Portugal e Espanha) progredindo de forma positiva e rápida. Já no início da década de 90 a situação modificou-se no Sul da Europa: a proporção de nascimentos fora do casamento aumentou rapidamente, e, exceptuando Portugal, todos os países atingiram valores inferiores a 1,3 filhos (Kohler et al. 2002).

No campo da fecundidade verificam-se diferentes comportamentos que se distinguem entre a primeira e a segunda transição demográfica. É importante uma vez mais reforçar que a primeira transição demográfica se caracteriza por deficientes métodos de contraceção assim como níveis de paridade social desiguais; sendo um período igualmente caracterizado por um declínio da fecundidade ilegítima; uma média de idades ao nascimento do primeiro filho mais baixa e um número reduzido de casais sem filhos. Em oposição na segunda transição demográfica, verifica-se um acentuado adiamento da fecundidade, e o aumento da média de idades da mãe no primeiro nascimento; métodos de contraceção eficientes em todos os estratos sociais; um aumento da fecundidade fora do casamento e também de casais sem filhos (Lesthaeghe e Neels 2002).

A rápida alteração do modelo de fecundidade, nomeadamente após a segunda Guerra Mundial, para níveis nunca anteriormente registados, levou muitos dos países europeus a experienciar os mais baixos níveis de fecundidade. De acordo com Frejka e Sobotka (2008), em 2002, 16 dos 39 países europeus com populações com mais de 100,000 habitantes (com excepção da Turquia) atingiram o valor de *ISF* de 1,3 e 25 países de 1,5. O adiamento da maternidade/parentalidade tem sido um factor crucial na evolução da fecundidade na última década, nomeadamente nos países desenvolvidos.

As constantes modificações nos padrões de fecundidade ampliaram a oportunidade de compreender estes comportamentos e o seu impacto na sociedade, principalmente em questões de interesse como a educação, a profissão e a situação profissional das mães, já que estas parecem ser as características fundamentais para o adiamento da fecundidade (Kohler et al. 2006). Nesta área destacam-se então investigadores como Sobotka (2003), Bongaarts e Feeney (1998), Andersson e Scott (2005), Kohler e Ortega (2002) entre outros, que ampliaram o campo de investigação da evolução da fecundidade.

Em pesquisas recentes Sobotka (2004) focou a sua análise na problemática dos nascimentos fora do casamento com e sem coabitação dos pais, assim como o aumento do espaçamento entre os nascimentos e o casamento, e o conseqüentemente crescente espaçamento entre os nascimentos. A sua análise, como a de outros autores (Kohler et al. 2006) também se tem centrado em torno de uma nova abordagem dos indicadores, através da utilização de *coortes* fictícias. Tal como outros autores (Kohler e Ortega ou Frejka *op. cit.*), também Sobotka (2004), por exemplo, realiza abordagens estatísticas na perspectiva demográfica, recorrendo à utilização de regressão linear múltipla, como acontece quando o autor tenta explicar a entrada de alguns países na segunda transição demográfica construindo dois índices que sintetizam algumas variáveis relacionadas, no primeiro com os comportamentos da população, e no segundo índice tenta-se explicar o enquadramento da SDT através da análise dos valores e comportamentos humanos.

Já Yavuz (2008) introduz uma nova abordagem, com base na Turquia, relativa à análise de fecundidade e à evolução do espaçamento entre os nascimentos dos filhos bem como para a evolução do número de filhos, propondo para a fecundidade a utilização de análise de sobrevivência, nomeadamente através de modelos de riscos proporcionais de Cox, conciliando a análise de sobrevivência com indicadores demográficos como o índice sintético de fecundidade.

A demografia e a estatística partilham “raízes” comuns, no entanto, estas disciplinas cresceram e desenvolveram-se separadamente. Quer a demografia, como a estatística, são ciências essenciais na análise de questões importantes da sociedade, como os sistemas de pensões, na estimação da dimensão populacional, ou na análise de problemas sociais como o desemprego. Ao propor novas abordagens estatísticas de análises demográficas, as duas ciências parecem caminhar para uma uniformização de ideias e métodos de análise, possibilitando não só análises mais adequadas, mas também mais precisas para explicar os comportamentos de uma dinâmica demográfica em constante alteração.

### **1.2.1. Impacto da Educação na fecundidade**

Com a difusão da pílula anticoncepcional, nos anos 70 do século XX, as mulheres e os casais passaram a ter ao seu alcance o instrumento que lhes faltava para controlar de forma bastante eficaz a sua fecundidade. Como referido anteriormente é nesta fase que a humanidade passa por uma segunda revolução da reprodução, que deixou de ser um acto inevitável, passando a ser uma opção do casal (Cunha, 2004).

Depressa a questão da educação no campo da fecundidade se tornou o factor que mais vezes surge como explicação, tanto no nível, como no espaçamento da fecundidade. Devemos ter em consideração que nas últimas décadas, os jovens europeus, têm vindo a “gastar” cada vez mais tempo da sua vida com os estudos, verificando-se um aumento de procura por uma educação mais elevada e flexível, levando ao adiamento da fecundidade (Sobotka, 2004). De facto, a escolaridade desempenha um papel decisivo no comportamento da fecundidade, tendo um papel incontornável nos estudos sobre os comportamentos procriativos e as representações no campo da demografia, nomeadamente, na fecundidade (Cunha, 2004).

A educação influencia indirectamente a evolução da fecundidade, as mulheres com níveis de educação mais elevados, tendo o seu primeiro filho mais tarde, rapidamente e muitas vezes por razões biológicas terão o segundo filho. Além disso, estas mulheres normalmente escolhem parceiros com níveis de educação semelhantes, a chamada “homogamia educacional”. Se tal se verificar, os homens com educação mais elevada devem usufruir de um rendimento mais elevado o que contribui para elevar o rendimento do agregado familiar (Caleiro, Mendes e Rego, 2006).

Existe em Portugal, de acordo com os dados do Inquérito à Fecundidade e à Família de 1997 (IFF), realizado pelo INE, uma inequívoca relação entre a idade da mulher ao nascimento do primeiro filho e o seu nível educacional; as mulheres com níveis educacionais superiores, apresentam idades ao nascimento do primeiro filho mais elevadas, confirmando assim que tendem a adiar o nascimento do primeiro filho em função directa do nível educacional (INE, 1998). Podemos então considerar que as mulheres com níveis de instrução mais baixos têm na maioria dos casos mais filhos, e o seu calendário inicia-se mais cedo do que o das mulheres com níveis de instrução mais elevados.

Alguns autores (Hoem, 2005; Andersson, 2005; Neyer, 2003) reportando-se à Suécia, consideram educação e fecundidade como um processo interactivo dinâmico, em que ambos se determinam mutuamente, na medida em que o sistema de ensino é flexível e as mulheres podem entrar e sair do sistema várias vezes ao longo da vida. Em Portugal, tal não acontece. O nível de educação atingido nas idades mais jovens será o determinante da fecundidade. No entanto, num futuro próximo e com o aumento da formação ao longo da vida, esta perspectiva de análise poderá também vir a alterar-se. Autores como os referidos anteriormente salientam ainda que a dinâmica entre educação e fecundidade pode ser mediatizada através de políticas públicas.

Assim, e apesar do leque de possibilidades que se abre hoje no domínio da parentalidade, tanto a nível económico como cultural, a maioria dos estudos europeus que reflectem esta nova realidade social têm referido uma grande homogeneidade das práticas e das representações: por um lado, o número de casais que optam por não ter filhos tem aumentado, e por outro, regista-se uma polarização em torno dos 2 filhos, sendo cada vez menos os casais que têm 3 ou mais filhos; por outro lado, é também importante reforçar a ideia de que se tem verificado um novo redimensionamento do modelo de descendência que passou de 2-4 filhos para apenas 2 (Cunha 2004).

Se associarmos o adiamento da fecundidade ao prolongamento da educação, então a verdade é que a situação portuguesa assim como nos restantes países da Europa torna-se apreensiva, visto que os jovens prolongam cada vez mais a sua educação, estendendo o seu tempo de estudo e encurtando simultaneamente o calendário da fecundidade, diminuindo também o número de filhos por casal e levando a um fraco (ou mesmo negativo) crescimento da população.

O aumento do número de mulheres com níveis educacionais elevados, inclusivamente no ensino superior, é uma característica da transformação que o seu papel na sociedade tem vindo a sofrer. Se no passado a mulher era sobretudo “dona de casa” e mãe, actualmente acrescenta às mesmas funções o papel de estudante e de trabalhadora activa. Os desejos de independência e de vencer numa sociedade que prima pela dinâmica individual são mais do que objectivos a atingir. Transformaram-se, essencialmente, numa forma de obter a subsistência a partir dos seus próprios meios em vez dos do homem, seja pai, irmão ou marido. A decisão do casal face à fecundidade depende de um vasto leque de condicionantes, dificilmente compreendidas pelos especialistas, que se crê derivarem em larga escala do meio económico, político e social

envolvente e das políticas natalistas utilizadas para favorecer o aumento da fecundidade (Silva 2002.).

### **1.2.2. Participação no Mercado de Trabalho**

São muitos os autores (Vikat 2004; Sobotka 2004; Brewster e Rindfuss 2000 entre outros) que colocam inúmeras questões sobre o impacto do mercado de trabalho e o papel da mulher em contraste com a fecundidade. Em função do grau de participação da mulher no mercado de trabalho são diversas as consequências ao nível da conciliação entre a vida profissional e a vida familiar, considerando-se até que poderá existir uma descendência populacional limitada por aqueles factores (Cunha 2004). No início da década de 80 entendia-se que trabalho e família seriam sempre, incompatíveis. Becker (1981, citado por Matsuo, 2003) colocou a hipótese de existir uma avaliação por parte dos casais, um balanço entre os factores económicos de ter uma família e o “preço” de substituir o valor económico das crianças que deixavam então de ser mão de obra lucrativa para se tornarem um encargo para os pais.

Além de considerarmos que a educação é uma condicionante da fecundidade também a relação com o mercado de trabalho condiciona a fecundidade e é por ela condicionada. Assim, e de acordo com alguns autores deste campo de análise o aumento da participação no mercado de trabalho por parte da mulher reflecte-se no declínio da fecundidade (Brewster e Rindfuss 2000). De acordo com Sobotka (2003.), no caso da participação da mulher no mercado de trabalho, tal como no caso da educação, o aumento de um factor parece ter implicação directa na diminuição do outro. Isto, significa que ao aumentar a participação no mercado de trabalho leva automaticamente à diminuição da fecundidade por parte da mulher, o que nem sempre se verifica (Mendes 1992). A força de trabalho feminina é enganadora no que respeita à questão de fundo sobre a fecundidade e suas alterações. Devido a esta situação, muitas nações industrializadas e desenvolvidas formularam políticas com base na associação destas duas características centrais na vida da mulher (Neyer 2003.). De certa forma a associação entre a fecundidade e o trabalho feminino parece reflectir a incompatibilidade entre educar uma criança e participar na vida económica e na produtividade das sociedades industrializadas (Brewster e Rindfuss 2000).

No campo socio-económico, as modificações nos padrões laborais europeus, nomeadamente no campo da participação feminina no trabalho, têm-se mantido como um factor predominante nas alterações de fecundidade na Europa. Desde a década de 70, que a participação feminina no mercado de trabalho tem vindo a aumentar significativamente principalmente nos países ocidentais do continente europeu. Na maioria dos países da Europa o trabalho feminino passou de cerca de 50% na década de 70 para cerca de 60% em meados da década de 90. No Sul da Europa (com excepção para Portugal onde a taxa de actividade feminina foi sempre elevada) a situação ainda que semelhante demorou mais tempo a atingir este nível de participação feminina (Neyer 2003).

Verifica-se aliás que a fecundidade pelo índice sintético de fecundidade na maioria dos países está abaixo do nível de substituição das gerações. Destacam-se países como a Grécia, Itália e a Espanha que têm desde 1995 os níveis mais baixos da Europa Ocidental (Silva 2002.). Embora variando de país para país, os comportamentos demográficos que determinam a fecundidade são influenciados por factores muito diversos que condicionam desde a nupcialidade ao momento do primeiro nascimento. Parece não restarem dúvidas de que as mulheres que desempenham funções mais elevadas no seu local de trabalho têm menos filhos quando comparadas com as outras, por outro lado, as mães com mais filhos ficam menos tempo nos seus empregos. A verdade é que a fecundidade é influenciada pelo trabalho feminino e vice-versa (Sobotka 2004).

Esta questão deverá então ser observada de dois pontos de vista, o primeiro partindo dos efeitos da fecundidade no trabalho e, no segundo caso, partindo dos efeitos do trabalho na fecundidade (Vikat 2004; Brewster e Rindufuss 2000). Contudo, na generalidade, podemos dizer que, tal como a fecundidade, o trabalho depende de múltiplas dimensões e factores que tornam estas questões tão complexas, mais do que pode parecer à primeira vista. Segundo Matsuo (2003) torna-se cada vez mais pertinente observarmos de perto os restantes factores da sociedade, como a cultura, os aspectos estruturais do sistema familiar e o próprio comportamento humano, levando-nos a uma compreensão mais profunda do problema levantando porém novas questões relacionadas com a educação dos filhos e a forma de criação das novas famílias.

A relação entre participação no mercado de trabalho e fecundidade também se prende com a falta de trabalho, já que na sua inexistência os filhos tendem a diminuir ou a

aumentar o adiamento da idade da mãe ao nascimento dos filhos. Existem ainda outras questões relacionadas com o trabalho em tempo parcial e com os empregos mais ou menos flexíveis, já que este tipo de empregos não oferece nem estabilidade nem segurança económica; por outro lado, é na sua grande maioria direccionado para a população feminina, mais disponível e com remunerações inferiores. Aliás muitas mulheres com empregos a tempo inteiro procuram ainda outras ocupações laborais que lhes permitam uma maior independência financeira perante as suas famílias e nomeadamente os seus companheiros (Sobotka 2003).

Embora ainda persista uma abordagem muito negativa sobre a relação entre o trabalho e a fecundidade, existem autores que defendem que relação é positiva. Aliás dizem-nos até que os factores educação e trabalho aumentam as expectativas familiares no que respeita ao casamento e aos filhos tornando-os o centro do esforço familiar, existindo uma correlação acentuada entre estes dois factores da vida adulta e familiar (Matsuo 2003).

## **2. Análise da fecundidade**

### **2.1. Dados**

Os dados utilizados na elaboração deste trabalho provêm de uma amostra dos micro-dados dos nascimentos relativos a Portugal para os anos 1995 a 2009 (cedidos pelo Instituto Nacional de Estatística<sup>2</sup>). Consideramos que os dados em análise são uma *mega amostra* na medida em que estamos a analisar todos os nascimentos ocorridos em Portugal, excluindo apenas deste estudo mãe de nacionalidade portuguesa que tiveram os seus filhos fora do país. Os dados em análise correspondem ao verbete preenchido no momento do nascimento de cada filho. Entre as diferentes variáveis disponíveis, decidiu-se analisar o ano de nascimento, o nível de instrução da mãe, a condição perante o trabalho da mãe, a profissão da mãe, a situação profissional da mãe, a idade da mãe e a filiação (posteriormente e no âmbito dos modelos estatísticos que nos propomos estabelecer foi contabilizado o tempo desde o casamento, bem como o tempo entre os nascimentos).

Foram ainda utilizados dados caracterizadores das mães, disponíveis no Human Fertilty Database<sup>3</sup>, para o mesmo período temporal, permitindo uma caracterização mais pormenorizada da situação demográfica em Portugal.

### **2.2. Análise descritiva**

Para podermos caracterizar a evolução da fecundidade em Portugal, ao longo do período de análise (de 1995 a 2009) e simultaneamente estudar, numa primeira abordagem, as variáveis que iremos utilizar ao longo deste trabalho, realizámos uma análise descritiva dos dados. A estatística descritiva pode definir-se como a área da estatística que visa sintetizar e apresentar os dados observados, através de quadros e gráficos que facilitem a interpretação dos dados (Maroco 2007).

Ao observarmos na figura 1 os nascimentos ocorridos em Portugal no período em análise, dentro e fora do casamento, verificamos que estes apresentam comportamentos diferentes. Entre 1995 e 2000 o número de nascimentos dentro do casamento apresentava um comportamento de crescimento, ainda que não acentuado, no entanto,

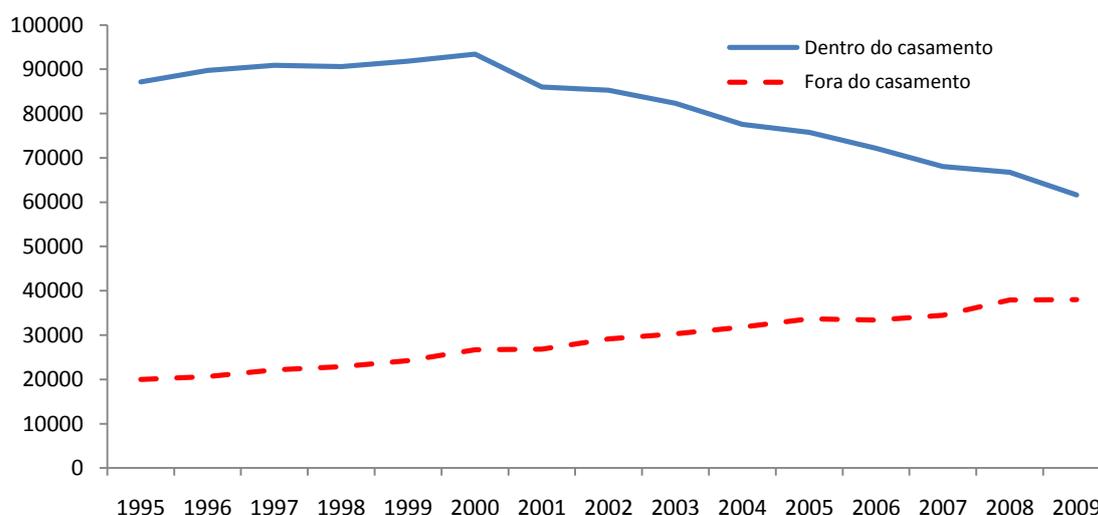
---

<sup>2</sup> <http://www.ine.pt>

<sup>3</sup> <http://www.humanfertility.org/cgi-bin/main.php>

após o ano de 2000, os nascimentos dentro do casamento apresentam um comportamento no sentido da diminuição. Já os nascimentos ocorridos fora do casamento apresentam um crescimento contínuo, desde 1995 até à actualidade. De destacar que a proporção de nascimento dentro e fora do casamento é ainda significativamente diferente.

**Figura 1:** Evolução dos nascimentos ocorridos em Portugal entre 1995 e 2009, dentro e fora do casamento



Relativamente à evolução dos nascimentos de acordo com o nível de escolaridade da mãe (tabela 1), verificamos que os comportamentos se têm vindo a modificar na medida em que se, por um lado, o número de mães sem escolaridade tem vindo a diminuir, por outro lado, o número de filhos de mãe com o ensino superior tem vindo a aumentar gradualmente. No entanto, é de destacar que um maior número dos nascimentos ocorre em mães que têm o ensino básico como nível de escolaridade.

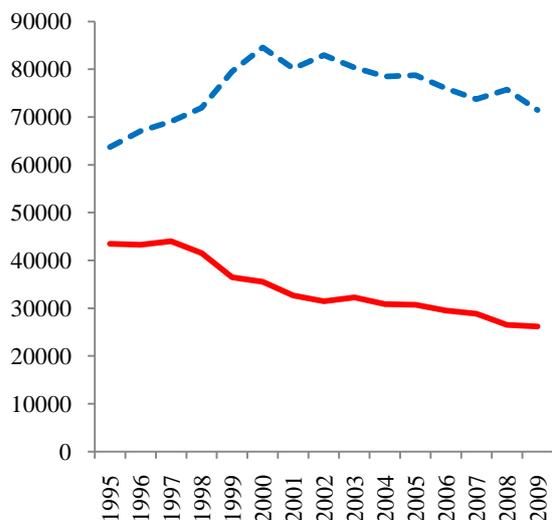
**Tabela 1:** Evolução dos nascimentos em Portugal entre 1995 e 2009 de acordo com o nível de escolaridade das mães

	Sem escolaridade	Ensino Básico	Ensino Secundário	Ensino Superior	Total
1995	3021	75221	17143	11786	107171
1996	2611	76227	18255	13252	110363
1997	2554	75178	20580	14720	113032
1998	2117	73098	22216	16069	113500
1999	2028	72644	23689	17668	116029
2000	1721	73213	25572	19554	120060
2001	1575	66923	24086	20235	112819
2002	1347	64304	26099	22696	114446
2003	1114	59530	27176	24767	112587
2004	1142	55946	26752	25513	109353
2005	1015	53315	27663	27461	109454
2006	870	49072	27365	27956	105510
2007	783	46317	26477	28319	102567
2008	646	45402	27833	29269	103150
2009	583	41187	27536	29335	98641
Total	23127	927577	368442	328600	1648682

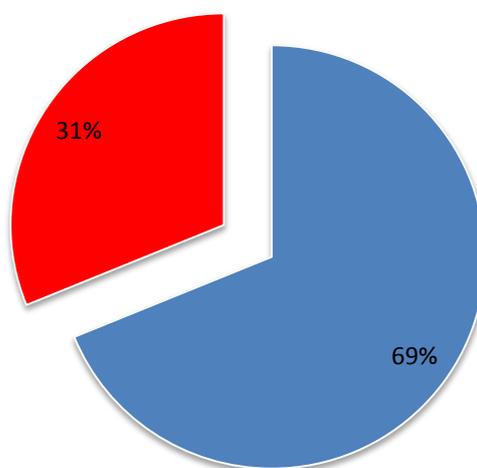
Relativamente à condição da mãe perante o trabalho verificamos na figura 3, e numa perspectiva geral do período em análise, que 69% das mulheres que foram mães estavam empregadas enquanto 31% se encontrava desempregada. Por outro lado, ao observarmos na figura 2 a evolução dos nascimentos entre 1995 e 2009 verificamos que tem vindo a diminuir o número de nascimentos de mães em situação de desemprego. Já nas mulheres empregadas que foram mães verifica-se um aumento até ao ano 2001 e, posteriormente a este ano, acontece também uma diminuição dos nascimentos.

**Figura 2 e 3:** Evolução dos nascimentos em Portugal de acordo com a condição perante o trabalho da mãe

**Figura 2:** Evolução entre 1995 e 2009



**Figura 3:** Comportamento Geral no período total em análise



Empregado(a)

Desempregado(a)

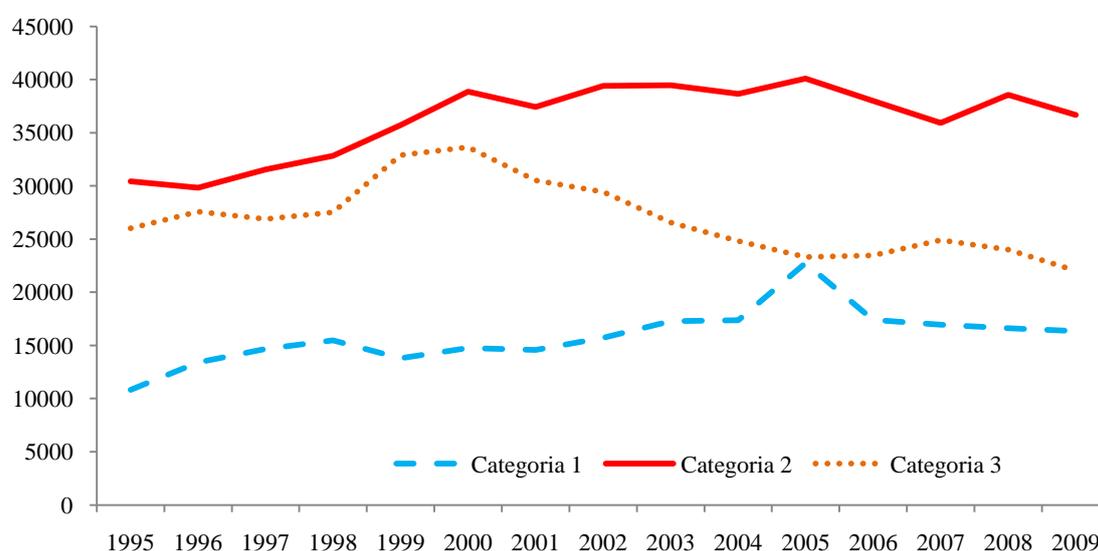
No que diz respeito à covariável profissão, é de referir que foi elaborada uma recodificação da covariável original, pois esta tinha 9 categorias o que tornava a sua utilização em alguns modelos estatísticos difícil. Assim, foi realizada a recodificação para apenas 3 categorias, que se apresenta na tabela 2. Esta reorganização das categorias profissionais foi realizada tendo em conta as semelhanças profissionais das mesmas categorias.

**Tabela 2:** Categorias Profissionais/Profissão

Categoria 1 (0)	Categoria 2 (1)	Categoria 3 (2)
Membros das forças armadas	Técnicos e profissionais de nível intermédio	Agricultores e trabalhadores qualificados da agricultura e pescas
Quadros superiores da administração pública, dirigentes e quadros superiores de empresas	Pessoal administrativo e similares	Operários, artífices e trabalhadores similares
Especialistas das profissões intelectuais e científicas	Pessoal dos serviços e vendedores	Operadores de instalações e máquinas e trabalhadores da montagem
-	-	Trabalhadores não qualificados

Quanto aos resultados obtidos com esta covariável (figura 4), verifica-se que um maior número de nascimentos ocorre nas mães pertencentes à categoria 2 onde se enquadram as categorias profissionais dos técnicos intermédios, o pessoal administrativo e dos serviços e os denominados vendedores. De destacar também que nas categorias extremas as situações são opostas, se na categoria 1 os nascimentos têm vindo a aumentar ligeiramente verificando-se um “pico” em 2005, já na categoria 3 os valores têm vindo a diminuir, reforçando os resultados obtidos ao analisarmos o nível de escolaridade da mãe, onde as mães com ensino superior aumentaram o número de filhos.

**Figura 4:** Evolução dos nascimentos em Portugal entre 1995 e 2009 de acordo com a Profissão das mães

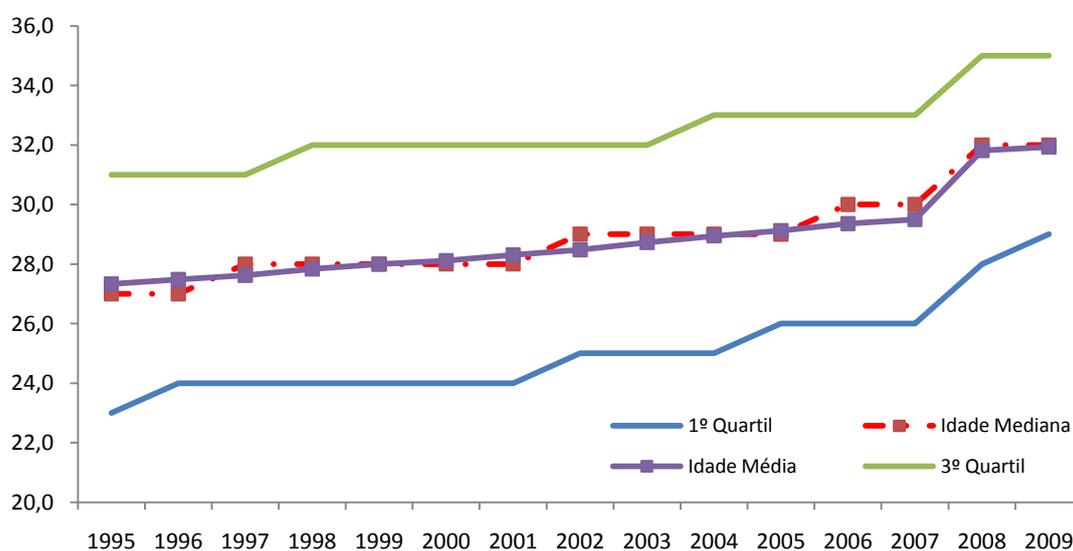


Na figura 5 podemos observar a evolução da idade mediana (segundo quartil) e média das mães ao nascimento do filho, bem como o primeiro e terceiro quartis. As medidas descritivas representadas permitem-nos observar um aumento da idade média a que as mães têm os seus filhos. É visível um aumento da idade das mães nomeadamente nos anos mais recentes (2008 e 2009). Podemos ainda verificar que a distribuição das idades das mães pela população em análise é bastante simétrica, uma vez que os valores obtidos para a mediana se encontram muito próximos dos valores medianos. Por outro lado, e apesar do destaque já dado aos dois últimos anos, importa referir que a alteração que se mostrou mais acentuada é referente ao 1º quartil, que em 2007 apresentava um valor de 26 anos de idade, para em 2009 atingir os 29. Isto significa que, se em 2007

25% das mães em Portugal registavam idades até 26 anos, em 2009 esse valor já era de 29 anos de idade.

Esta situação ainda se clarifica mais, se forem analisadas as evoluções dos 3 quartis desde 1995, onde 25% das mães registavam idades até 23 anos, 50% até 27 anos, e 75% destas registavam até 31 anos. Em contrapartida, no ano de 2009 as idades registaram um aumento significativo, apresentando valores de 29, 32, e 35 anos de idade, para o 1º, 2º, 3º quartis, respectivamente. Denota-se assim, um claro e evidente deslocamento da idade ao nascimento dos filhos, das mães portuguesas, para idades mais avançadas.

**Figura 5:** Evolução da idade das mães em Portugal entre 1995 e 2009



Na tabela 3 podemos observar nos diferentes anos do período em análise algumas medidas de dispersão que complementam a figura 5. Nesta tabela observa-se a média, bem como os seus intervalos de confiança, a média aparada a 5% a mediana, a variância, o desvio padrão, bem como os valores mínimo e máximo registados nos diferentes anos e a sua amplitude e por fim as medidas de achatamento e de enviesamento. Observa-se assim um aumento da idade média ao nascimento dos filhos para as mulheres que com filhos no período entre 1995 e 2009. No que diz respeito aos valores mínimos da idade de mulheres que foram mães neste período observa-se um padrão que varia entre os 12 e 13 anos sendo apenas diferente nos anos de 2008 e 2009 com idades de 16 e 17 como valores mínimos. Já na idade máxima a que as mulheres foram mães no período em análise, verifica-se que este valor variou entre os 52 e 54

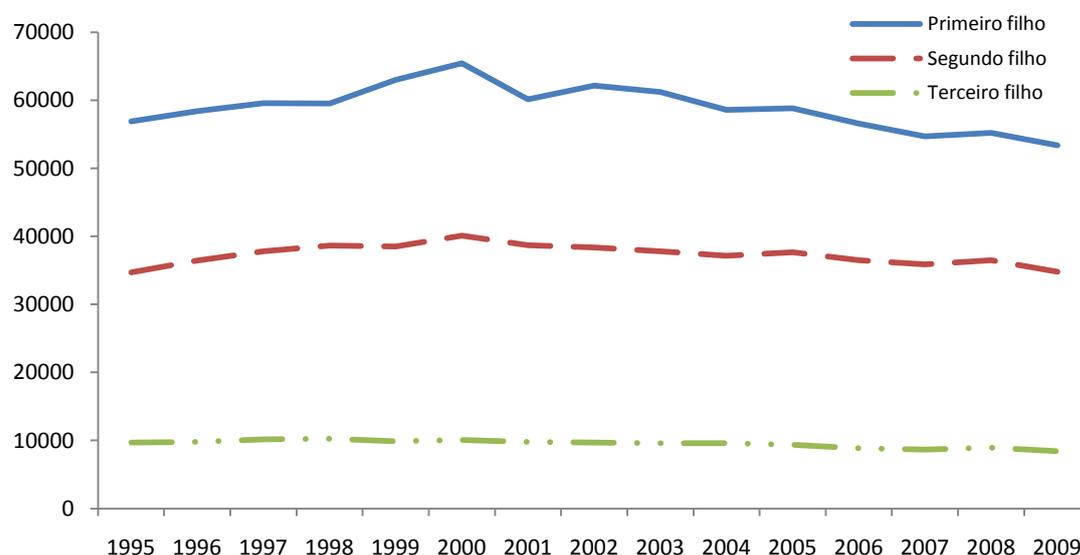
anos para todos os anos com excepção uma vez mais de 2008 e 2009, onde a idade mais tardia a que as mães tiveram os seus filhos foi aos 47 anos, verificando-se uma diminuição na idade máxima. Desta forma também a amplitude entre as idades apresentou também uma diminuição de 40 anos em 1995 para os 30 anos em 2009, embora esta amplitude tenha diminuído 10 anos entre 1995 e 2009, isso não diminui a dificuldade de construir modelos onde esta covariável apareça como continua dado a complexidade dos comportamentos bem como a dinâmica própria de cada idade, por outro lado no campo da demografia é muitas vezes realizada uma análise da idade da mãe em grupo de idade quinquenais. Mais à frente serão apresentadas diferentes codificações para a idade da mãe de forma a compreender o fenómeno da evolução do espaçamento entre os nascimentos e os factores que influenciam este comportamento.

**Tabela 3:** Medidas de dispersão para a covariável Idade da mãe no período em análise

	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	
Média	27,3	27,5	27,6	27,8	28,0	28,1	28,3	28,5	28,7	28,9	29,1	29,4	29,5	31,8	31,9	
95% I.C. Média	Lim. Sup.	27,3	27,4	27,6	27,8	28,0	28,1	28,3	28,4	28,7	28,9	29,1	29,3	29,5	31,7	31,9
	Lim. Inf.	27,4	27,5	27,7	27,9	28,0	28,1	28,3	28,5	28,8	29,0	29,2	29,4	29,5	31,9	32,0
Média aparada 5%	27,2	27,4	27,5	27,8	27,9	28,1	28,3	28,5	28,7	28,9	29,1	29,4	29,5	31,8	32,0	
Mediana	27,0	27,0	28,0	28,0	28,0	28,0	28,0	29,0	29,0	29,0	29,0	30,0	30,0	32,0	32,0	
Variância	29,9	29,8	30,1	30,1	30,6	31,0	31,0	30,9	30,6	31,0	30,6	30,5	30,7	23,9	24,1	
Desvio Padrão	5,5	5,5	5,5	5,5	5,5	5,6	5,6	5,6	5,5	5,6	5,5	5,5	5,5	4,9	4,9	
Mínimo	13	12	12	12	12	12	12	12	13	13	13	12	13	16	17	
Máximo	53	52	54	53	53	53	52	54	53	54	53	52	54	47	47	
Amplitude	40	40	42	41	41	41	40	42	40	41	40	40	41	31	30	
Enviesamento – “Skewness”	0,3	0,2	0,2	0,1	0,1	0,1	0,1	0,1	0,0	0,0	-0,1	-0,1	-0,2	-0,1	-0,1	
Achatamento – “Kurtosis”	-0,2	-0,3	-0,3	-0,3	-0,3	-0,3	-0,3	-0,3	-0,3	-0,2	-0,2	-0,2	-0,2	-0,3	-0,2	

Relativamente à evolução do número de filhos, entre 1995 e 2009 (figura 6), podemos observar que o número de mães com o seu primeiro filho tem vindo a diminuir, verificando-se, no entanto, um ligeiro aumento no ano de 2000, considerados por muitos, um mini “baby boom”. O comportamento relativo à evolução do número de segundos filhos, e à semelhança do que acontece no primeiro filho, apresenta um ligeiro aumento no ano de 2000, seguido de um declínio.

**Figura 6:** Evolução do número de nascimentos de acordo com a ordem dos mesmos em Portugal entre 1995 e 2009



Realizada uma breve abordagem de estatística descritiva das variáveis em análise é também importante uma segunda abordagem que passa por analisar os indicadores demográficos de fecundidade, como o Índice sintético de Fecundidade e a Idade média à Fecundidade, com o objectivo de completar a análise descritiva das variáveis em estudo.

### 2.3. Indicadores demográficos

É importante, antes de mais, compreender que qualquer fenómeno demográfico pode ser analisado segundo duas perspectivas distintas: a primeira, a óptica longitudinal, segue o percurso de vida das gerações, enquanto a segunda perspectiva, a análise transversal, resulta da observação de múltiplas gerações num só momento do tempo (Oliveira, 2008). Esta questão é importante na medida em que é através destas perspectivas que devemos direccionar a análise demográfica, nomeadamente pela dificuldade de obter informação que permita elaborar os dois tipos de abordagem.

### 2.3.1. Metodologia

Assim, e considerando a análise transversal ou por *período* como aquela que se revela mais pertinente, pois a análise da fecundidade através desta metodologia é mais acessível e os resultados obtidos são imediatos, poderemos assim obter o Índice Sintético de Fecundidade (*ISF*). Ao utilizar o *ISF* estamos a calcular o número de filhos que em média uma mãe deixa naquele momento se as condições se mantiverem estáveis. O *ISF* obtém-se através de:

$$ISF = (\sum TFG_i) * 5 \quad [2.1]$$

e onde a,  $TFG_i$  (taxa de fecundidade geral por idades) =  $\frac{\text{Total de nados vivos por idades das mães}}{\text{População feminina idade a idade}}$ .

O *ISF* é uma medida calculada em transversal, ou seja, calculada num determinado momento do tempo, normalmente num ano civil, e que mede o número de filhos que em média uma mulher deixa na população, se mantiver ao longo de toda a sua vida fértil um comportamento idêntico ao registado nesse mesmo ano civil no conjunto das mulheres existentes na população, com idades entre os 15 e os 50 anos exactos e que, pertencem assim a diferentes gerações (Mendes e Rego 2007). Por outro lado, e tendo em conta que este índice é fortemente influenciado pela idade em que as mulheres decidem ter filhos, faz todo o sentido analisar, conjuntamente, a Idade Média da Fecundidade (*IMF*), quer no seu todo, quer por ordem de nascimento:

$$IMF = \frac{(\sum TFG_i * x_i)}{\sum TFG_i} \quad [2.2]$$

Sendo o *ISF* a medida mais utilizada para analisar a fecundidade e também o indicador mais representativo do período fecundo, é, simultaneamente, uma medida mais susceptível de ser influenciada pelo efeito tempo (Sobotka, 2004). Assim, seguindo Bongaarts e Feeney (1998), calculámos o *ISF* ajustado, neutralizando o efeito ao tempo, no indicador do momento, provocado pelo adiamento observado na *IMF*. Este indicador é calculado através de:

$$adjISF_i(t) = \frac{ISF_i(t)}{[1-r_i(t)]} \quad [2.3]$$

em que o  $r$  é o valor médio da diferença entre a idade média das mães ao nascimento do filho no ano imediatamente anterior ( $t - 1$ ) e no ano seguinte ( $t + 1$ ) ao ano em análise ( $t$ ).

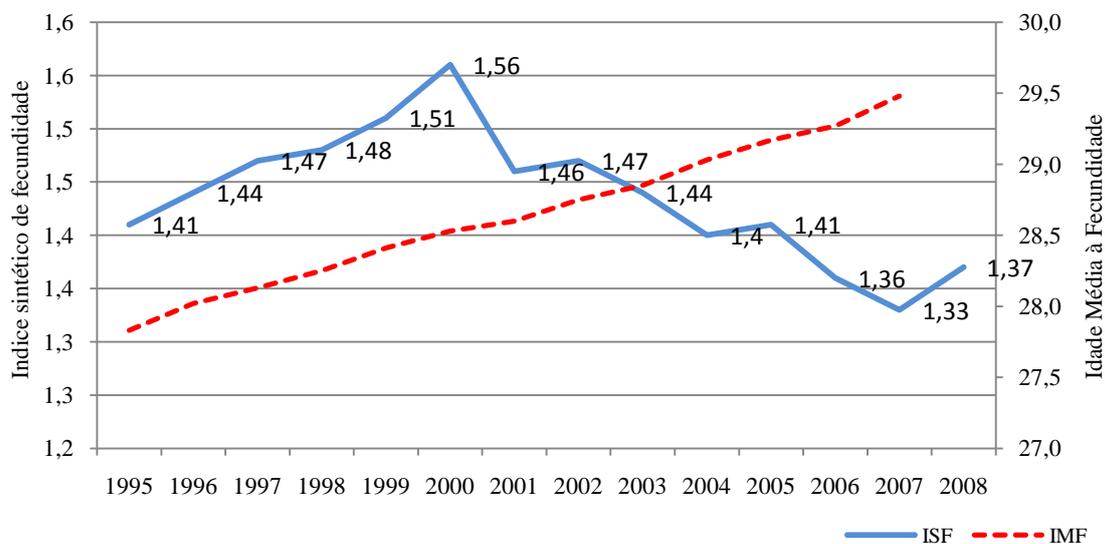
Desta forma isolamos o efeito do tempo no cálculo do  $ISF$ , na medida que se trata, como já referimos, de um indicador da fecundidade do momento. No caso de se verificar um aumento da idade média da fecundidade ao longo do tempo, o  $ISF$  tende a subavaliar a fecundidade. No caso contrário, isto é, se a idade média diminuir, tende a sobrevalorizá-la.

O nível de reposição da população assenta nos 2,1 filhos por casal, para um nível de mortalidade idêntico ao registado em Portugal, nos nossos dias, sendo, deste modo, importante compreender de seguida a evolução portuguesa do índice sintético de fecundidade bem como da idade média da fecundidade.

### 2.3.2. Resultados

Observando as tendências da fecundidade (Figura 7) através do próprio Índice Sintético de Fecundidade ( $ISF$ ), e da Idade Média à Fecundidade ( $IMF$ ), verificamos que à medida que o  $ISF$  diminui a  $IMF$  aumenta, ou seja, diminui o número de filhos que em média uma mulher deixa na população enquanto aumenta a idade média em que as mulheres têm os seus filhos. Observamos também que, no ano de 2000 o  $ISF$  é o mais elevado do referido período, com o valor de 1,56, confirmando um maior número de nascimentos ocorridos nesse ano. Embora o  $ISF$  tenha aumentado relativamente entre 1995 e 2000, a verdade é que nos últimos anos ele tem vindo a diminuir atingindo, em 2007, o valor mais baixo com 1,33 filhos por mulher. A gradual diminuição do  $ISF$  é acompanhada substancialmente pelo adiamento da fecundidade.

**Figura 7:** Evolução do Índice Sintético de Fecundidade e Idade Média à Fecundidade

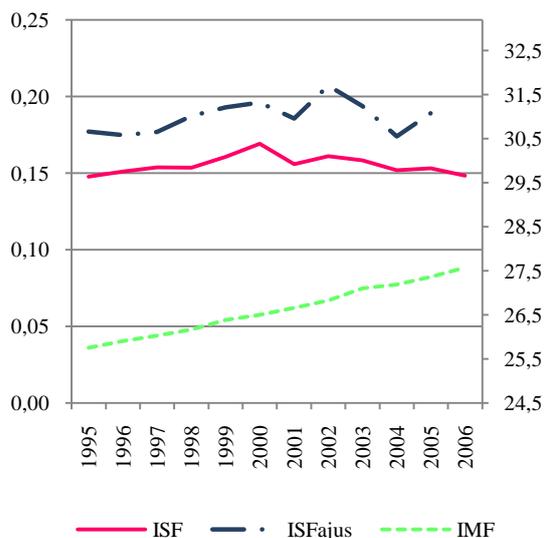


Embora seja de extrema importância a observação do ISF e da IMF no contexto geral dos nascimentos, é também importante compreender a evolução destes indicadores de acordo com a ordem do nascimento, nomeadamente no caso do primeiro, segundo, e terceiro filho. Por outro lado é também importante observar o impacto do ISF corrigido para efeito do *tempo* e do *quantum*.

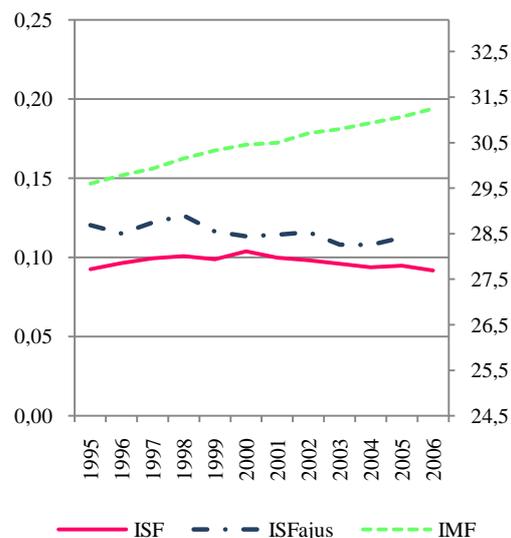
Na figura 8, e no que diz respeito ao primeiro filho, observa-se uma tendência crescente a IMF, em contraposição com uma ligeira diminuição do ISF, nomeadamente do *adjISF*. No caso do segundo (figura 9) e do terceiro filho (figura 10) verifica-se uma situação semelhante com a IMF a aumentar entre 1995 e 2006, enquanto o ISF se mantém constante mas a declinar. De uma forma geral os comportamentos do ISF e do IMF são semelhantes, independentemente da ordem dos nascimentos, pois a realidade é que à medida que um aumenta (IMF) o outro diminui (ISF), e nem mesmo o *adjISF* parece apresentar grandes melhorias nos resultados apresentados, ou seja, ainda que as mulheres deixassem de adiar a sua fecundidade esta não iria aumentar.

**Figura 8 a 10:** Evolução do *ISF*, *IMF* e *adjISF* em Portugal entre 1995 e 2006

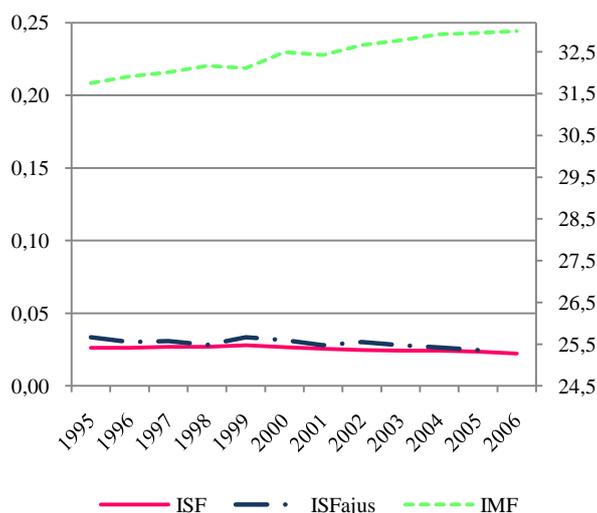
**Figura 8:** 1º Filho



**Figura 9:** 2º Filho



**Figura 10:** 3ºFilho



Os resultados que os indicadores demográficos nos apresentam não são muitas vezes suficientes para compreender e analisar os comportamentos das populações. Assim torna-se cada vez mais necessário analisar os dados de que dispomos com novas/diferentes metodologias, nomeadamente metodologias estatísticas. Verificámos que o número de nascimentos tem vindo a diminuir nesta última década, no entanto, é também importante verificar se o espaçamento entre os diferente eventos relacionados com o nascimento de filhos se tem mantido ou se, pelo contrario, se tem alterado. Assim, e com o objectivo de analisar os diferentes espaçamentos entre o casamento e o nascimento do primeiro filho, entre o primeiro e o segundo filho e, por fim, entre o

segundo e o terceiro filho, iremos no próximo capítulo recorrer à Análise de Sobrevivência.

### 3. Análise de Sobrevivência

A análise de sobrevivência é uma metodologia estatística, utilizada em diferentes contextos que se interessam pela ocorrência de *eventos*, onde a covariável resposta é o tempo até à ocorrência do evento. Os *eventos* deverão ser entendidos como sendo os acontecimentos na vida dos indivíduos e o risco de estes acontecimentos se concretizarem. A análise de sobrevivência é utilizada do ponto de vista de diferentes ciências como a medicina, sociologia, biologia e a demografia, sendo que não é uma análise restringida aos humanos, podendo ser realizada sobre animais ou equipamentos mecânicos ou de outro tipo. No campo da demografia a análise de sobrevivência está muito associada à mortalidade, no entanto, a sua metodologia poderá ser utilizada para analisar migrações, divórcio, casamento e fecundidade. Este tipo de análise tem duas características fundamentais: (1) devido à sua assimetria positiva que habitualmente este dados assumem, não permite que se assumam uma distribuição normal; (2) por outro lado, e devido às próprias características dos dados em algumas situações, a análise assenta em dados censurados. Os dados censurados acontecem quando, no momento do estudo, o evento ainda não aconteceu para alguns dos indivíduos, ou, por exemplo, quando por algum motivo alheio ao próprio estudo os indivíduos o abandonam.

#### 3.1. Metodologia

Considere-se  $T$  uma covariável aleatória (*v.a.*) que representa o tempo de vida de um indivíduo numa dada população; sendo  $F(t)$  a função de distribuição (*f.d.*) de  $T$  e definida como a probabilidade de um indivíduo experienciar determinado evento no intervalo  $[0, t]$ :

$$F(t) = P(T \leq t), \quad t \geq 0 \quad [3.1]$$

onde o  $P$  representa a probabilidade de o evento ocorrer.

Em análise de sobrevivência,  $S(t)$  a função de sobrevivência (*f.s.*) de  $T$ , é definida como a probabilidade de um indivíduo sobreviver além do instante  $t$ :

$$S(t) = 1 - F(t) = P(T > t), \quad t \geq 0 \quad [3.2]$$

Outro conceito importante em análise de sobrevivência é a função hazard ( $f.h.$ ). Esta função, e dependendo do âmbito da sua aplicação, é também conhecida como taxa de mortalidade, taxa de incidência, curva de mortalidade ou força da mortalidade. Desta forma a função hazard, representando a taxa instantânea até ao evento to instante  $t$  e condicionada à sobrevivência até esse instante, é definida como:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T > t)}{\Delta t} = \frac{f(t)}{S(t)}, \quad t \geq 0 \quad [3.3]$$

onde se assume que o  $T$  é uma *v. a.* contínua.

A função hazard pode levar a conclusões sobre a evolução de padrões de ocorrência, sendo ainda mais informativa que a função de sobrevivência. A representação gráfica desta função assume algumas tipologias, desta forma, a função de risco pode ser crescente (por exemplo, o envelhecimento gradual de algo), crescente (por exemplo, no caso de uma doença congénita), constante (com uma distribuição exponencial, por exemplo, em casos de acidentes ou doenças raras) e, por fim, temos a função com forma “bathtub shaped”, sendo decrescente, constante e depois crescente no final (o comportamento de qualquer população desde que nasce até ocorrer a morte).

É ainda importante definir, a função hazard cumulativa ou hazard integrada:

$$H(t) = \int_0^t h(u) du \quad [3.4]$$

De destacar que entre as funções  $f(t)$ ,  $h(t)$  e  $S(t)$ , podem ser estabelecidas relações:

$$h(t) = \frac{f(t)}{S(t)} = \frac{S'(t)}{S(t)} = \frac{d}{dt} \ln[S(t)] \quad [3.5]$$

Por outro lado e sabendo que  $S(0) = 1$ , temos então que:

$$S(t) = \exp[-H(t)] \quad [3.6]$$

e da mesma forma temos assim que:

$$H(t) = -\ln S(t) \quad [3.7]$$

### 3.1.1. Censura & Truncatura

A censura é a característica principal que diferencia a análise de sobrevivências de outras metodologias estatísticas. De forma genérica, uma observação censurada é uma observação incompleta que contém apenas informação parcial sobre um evento. No caso de se verificar a existência de dados censurados, isso acontece porque existem indivíduos para os quais não foi possível observar o seu tempo até ao evento com exactidão, havendo assim uma informação incompleta. Já a truncatura, ocorre quando se pretende apenas realizar uma selecção dos casos a estudar. Quer na censura como na truncatura existem diferentes padrões.

No caso da censura temos 4 padrões diferentes, (1) censura à direita; (2) à esquerda; (3) intervalar; (4) aleatória. A censura à direita acontece quando apenas se sabe que o tempo de vida excede um determinado valor, pois o indivíduo objecto de estudo o abandonou antes que ocorra o evento de interesse. Já a censura à esquerda ocorre quando o evento de interesse surge mas o estudo continua a decorrer independentemente disso, sendo que o evento pode até ter ocorrido antes de o indivíduo ter entrado no estudo. Já a censura intervalar, como o próprio nome indica ocorre num determinado intervalo de tempo, sem que se saiba o momento exacto em que o evento de interesse ocorreu. A censura aleatória acontece quando o tempo de observação é uma observação é uma covariável aleatória e os tempos de censura (potenciais) são variáveis aleatórias mutuamente independentes e independentes dos tempos de vida.

No caso da truncatura existem dois tipos, (1) a truncatura à esquerda e (2) truncatura à direita. A truncatura à esquerda ocorre quando apenas são incluídos no estudo indivíduos que satisfazem determinada condição que deve ocorrer antes do evento do interesse do estudo. Já na truncatura à direita apenas se observa o indivíduo se o acontecimento de interesse tiver ocorrido antes de momento específico, isto é, o tempo de sobrevivência deverá ser menor a um valor determinado pelo investigador.

Os estudos de análise de sobrevivência envolvem muitas vezes covariáveis que estão relacionadas com o tempo de sobrevivência. Estas covariáveis devem ser incluídas na

análise estatística dos dados para explicar o seu possível efeito no tempo de sobrevivência de um dado indivíduo.

O modelo de riscos proporcionais é uma das alternativas metodológicas que incorpora informações no estudo do tempo de sobrevivência através da introdução de covariáveis (é o modelo de riscos proporcionais). A família de riscos proporcionais é uma classe de modelos com a propriedade em que diferentes indivíduos têm funções de riscos proporcionais. Ou seja, a razão entre duas funções de riscos para dois indivíduos distintos não varia com o tempo (Hosmer e Lemeshow, 1999).

Compreendidos o tipo de dados que se encontram associados à análise de sobrevivência, é também importante compreender qual o modelo de estimativa adequado ao tipo de análise referido e aos dados disponíveis. Desta forma passamos a explicar em seguida o modelo de Cox bem como o modelo aditivo de Aalen.

### 3.1.2. Modelo de Cox

Sob a suposição de riscos proporcionais, Cox propôs em 1972 o Modelo de Riscos Proporcionais de Cox (Hosmer e Lemeshow, 1999). Este modelo é a técnica estatística utilizada fundamentalmente em análise de sobrevivência sendo um modelo semi-paramétrico baseado na função de hazard não assumindo nenhuma distribuição é o modelo semi-paramétrico mais utilizado. É um modelo estatístico de regressão reconhecido pela sua versatilidade e flexibilidade. O modelo de Cox é caracterizado pelos seus coeficientes  $\beta$  que medem os efeitos das covariáveis sobre a função de taxa de risco. Apresentando-se de uma forma geral como:

$$h(t, X) = h_0(t) \exp\{\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p\} \quad [3.8]$$

em que  $h_0(t)$  é o risco base no instante de tempo  $t$ .

O  $h_0(t)$  é a componente não paramétrica deste modelo, que apresenta como componente paramétrica os coeficientes  $\beta$  que medem o efeito das covariáveis sobre a função de risco. A regressão de Cox é considerada um modelo semi-paramétrico, porque a função de hazard ( $h(t)$ ) e a probabilidade de distribuição do tempo de

sobrevivência, não têm de ser especificados. Uma vez que a *baseline hazard* não é especificada, é utilizado um diferente parâmetro para cada tempo de sobrevivência. Este modelo permite-nos observar a associação entre o tempo de sobrevivência e uma ou mais variáveis independentes (binárias ou contínuas).

O modelo de riscos proporcionais de Cox pressupõe duas hipóteses: (1) assume que, por exemplo, a taxa de risco de dois indivíduos é independente do tempo e sendo este pressuposto válido apenas para as covariáveis independentes do tempo - isto significa que as funções de hazard para dois indivíduos em qualquer momento do tempo são proporcionais; (2) assume-se no modelo de Cox que a censura deve ser independente de qualquer incidente.

É, no entanto, importante ter em conta que para a utilização do modelo de riscos proporcionais de Cox é necessário verificar se de facto este modelo é um modelo de riscos proporcionais. Assim é necessário verificar de forma matemática e/ou gráfica se os pressupostos da proporcionalidade se verificam. O objectivo dos pressupostos pode desta forma ser a verificação (1) da proporcionalidade, ou seja, a relação entre a covariável resposta e a covariável independente do tempo, (2) ou a verificação da log-linearidade (já que a função tem uma estrutura log-linear), isto é, por exemplo, a **razão de riscos** entre um indivíduo de 30 anos e outro de 35 é semelhante àquela verificada entre um indivíduo de 90 e 95 anos.

Desta forma e para verificar a adequabilidade do modelo temos os resíduos de Schoenfeld padronizados que nos permitem avaliar o pressuposto dos riscos proporcionais, após o ajustamento de um modelo de Cox:

$$r_i(\beta) = x_i - \frac{\sum_{j \in R(t_i)} x_j \exp(x, \beta)}{\sum_{j \in R(t_i)} \exp(x, \beta)} \quad [3.9]$$

onde  $j$  representa cada indivíduo e  $i$  ( $i = 1, \dots, m$ ) é o índice dos tempos observados de eventos. Os resíduos de Schoenfeld são a diferença entre os valores observados de covariáveis para um individuo com tempo de ocorrência do evento  $t_i$  e os valores esperados em  $t_i$  dado o grupo de risco  $R(t_i)$ . Por outro lado e assumindo um coeficiente  $\beta_k$  (sendo que  $k$  é cada covariável) que varia com o tempo  $t$ , e em que  $\beta_k$ , poderá ser

uma média constante ( $E[r_i(\beta_k)|R(t_i)]$ ), com uma variância  $V(\beta_k)$  e uma função  $U(t)$  que varia com o tempo, os resíduos padronizados de Schoenfeld podem ser obtidos por:

$$r_i * (\beta) = \frac{r_i(\beta_k)}{V(\beta_k)} \quad [3.10]$$

O valor esperado deste resíduo padronizado  $r_i * (\beta)$  para cada grupo em risco  $R(t_i)$  é aproximadamente igual à parte de  $\beta_k$  que varia no tempo.

Por um lado podemos observar os resíduos martingala, que apresentam a sua utilidade na verificação da forma funcional de cada covariável, de modo a explicar o melhor possível o seu efeito na sobrevivência e sendo também utilizado para a detecção de *outliers*. Os resíduos martingala, são a diferença entre o número observado de eventos para um indivíduo e o esperado, dado o modelo previamente ajustado, o tempo e percursos observados de quaisquer covariáveis dependentes do tempo. Estes resíduos são analisados graficamente, onde podemos ter, o  $M_i$  versus o índice de cada indivíduo permitindo revelar os indivíduos mal ajustados pelo modelo, ou  $M_i$  do modelo nulo versus covariáveis com a sobreposição de uma curva de alisamento, para avaliar a forma funcional da covariável a ser incluída no modelo.

É ainda importante referir, os resíduos *Score*, que verificam a influência de cada observação no ajuste do modelo, permitindo também a estimação robusta da variância dos coeficientes de regressão.

Apesar de todos os pressupostos para verificar os riscos proporcionais associados ao modelo de Cox, aquando de estudos em populações e não em amostras, nem sempre a proporcionalidade dos riscos é assumida devendo esta questão ser tida em conta. De acordo com Rocha e Papoila (2009) é importante compreender se a não proporcionalidade é de facto importante, no entanto, esta questão não será abordada neste trabalho, considerando-se assim que será necessário verificar a existência destes pressupostos.

### 3.1.3. Estimador de Kaplan-Meier

O estimador Kaplan-Meier (KM) é o mais utilizado em estudos estatísticos que utilizam análise de sobrevivência, conhecido também como estimador “product-limit” (produto limite). Este é um estimador não paramétrico para a função de sobrevivência, e a sua construção considera o número de intervalos iguais ao número de falhas distintas e os limites dos intervalos são os próprios tempos de falhas da amostra. Este estimador não paramétrico para a função de sobrevivência é dado por,

$$\hat{S}(t) = \prod_{i:t_i \leq t} \left(1 - \frac{d_i}{n_i}\right) \quad [3.11]$$

onde  $0 \leq t \leq t_1$  para  $i=1$  e  $t \geq t_1$  para  $i \geq 2$   $\prod_{i:t_i \leq t} \left(1 - \frac{d_i}{n_i}\right)$ .

A estimativa de  $\hat{S}(t)$  pode se obtida por:

$$\hat{S}(t) = 1 \quad 0 \leq t \leq t_1$$

$$\hat{S}(t_1) = \hat{S}(t_1^+) = \frac{n_1 - d_1}{n_1}$$

$$\hat{S}(t_i) = \hat{S}(t_{i-1}) \times \frac{n_i - d_i}{n_i} \quad i=1, \dots, k$$

Sendo que a função  $\hat{S}(t)$  é uma função em escala que decresce após cada instante de ocorrência do evento. Os saltos da função depende não só do número de eventos, mas também do número de indivíduos em risco no momento são expressos por um factor  $\frac{n_i - d_i}{n_i}$ . De destacar também que se a maior observação registada for não censurada, então  $\hat{S}(t)$  toma o valor zero a partir desse instante. No caso em que a maior observação registada seja um tempo de censura, considera-se que  $\hat{S}(t)$  está definida apenas até este instante, nunca atingindo o valor zero.

### 3.1.4. Modelo Aditivo de Aalen

Embora o modelo de riscos proporcionais de Cox apresente a vantagem de uma acessível e simples interpretação, nomeadamente ao adicionar novas covariáveis, e ser facilmente estendido para covariáveis dependentes do tempo, Aalen aponta duas limitações ou problemas no modelo de Cox. A primeira limitação apontada por Aalen, está relacionada com situações onde a utilização do modelo de Cox é realizada sem a verificação dos seus pressupostos. Uma segunda limitação identificada, faz referência ao facto de o modelo de Cox não ser para detectar mudanças de efeitos de covariáveis ao longo do tempo. E é ainda importante referir que a suposição da proporcionalidade dos modelos é vulnerável às mudanças ocorridas no número de covariáveis modeladas (Colosimo e Giolo, 2006).

Estas limitações apontadas por Aalen e referidas por Colosimo e Giolo (2006), conduziram a modelos alternativos ao modelo de Cox. Em 1980 Aalen propõe um modelo de riscos para a análise de regressão de dados censurados como alternativa ao modelo de riscos proporcionais de Cox. O modelo de Aalen permite assim que os parâmetros das co-variáveis variem com o tempo, sendo o modelo capaz de fornecer informações detalhadas relativamente à influência temporal de cada co-variável. Desta forma, e contrariamente ao modelo de Cox, o modelo de Aalen é completamente não-paramétrico sendo que são as funções que são ajustadas e não os parâmetros.

Como referido anteriormente, no modelo de riscos proporcionais de Cox, os coeficientes de regressão  $\beta$ , representam os efeitos deste modelo, sendo constantes desconhecidas e cujos seus valores não mudam com o tempo. Já no modelo aditivo de Aalen assume-se que as co-variáveis actuam de forma aditiva na função de risco base, e as funções de regressão, ou coeficientes de risco ( $\beta(t)$ ) são funções do tempo, podendo o efeito das co-variáveis variar durante o estudo.

Desta forma e seja  $x = (1, x_1, x_2, \dots, x_p)'$ , o conjunto de valores das co-variáveis no tempo zero, o estimador do risco acumulado é:

$$\widehat{\Lambda}(t) = x'\widehat{\beta}(t) = \widehat{\beta}_0(t) + \sum_{j=1}^n \widehat{\beta}_j(t) x_{ij}(t) \quad [3.12]$$

onde  $t \leq r$ , e assumindo  $\widehat{\beta}_j(t)$  como os estimadores dos mínimos quadrados.

Por outro lado e tendo em conta a relação já referida anteriormente entre a função de sobrevivência e a função de risco acumulado, a sobrevivência ( $S(t)$ ), é dada por:

$$S(t) = \exp\{-\widehat{\Lambda}(t)\} \quad [3.13]$$

Como alternativa e baseado no estimador de Kaplan-Meier, a função de sobrevivência pode também ser estimada de acordo com a equação 3.14, não sendo necessariamente monótona sobre o período de observação.

$$\widehat{S}(t) = \prod_{t_i \leq t} [1 - (Z(t_i)I_k)'X] \quad [3.14]$$

onde  $I_k$  é um vector de zeros que assume o valor 1 para um dado individuo cujo evento ocorre no tempo  $t_i$  e  $Z(t)$  é a inversa de  $X(t)$ , isto é,

$$Z(t) = [X(t)'X(t)]^{-1} X(t)' \quad [3.15]$$

Aquando da utilização do modelo aditivo de Aalen é também importante avaliar se uma covariável tem algum efeito na função de risco total, testando-se uma hipótese nula, em que não existe efeito da co-variável sobre a função de risco, sendo a hipótese estabelecida:

$$H_{0j}: \beta_j(t) = 0, \quad t \in [0, r] \quad [3.16]$$

O modelo aditivo de Aalen apresenta-se como uma alternativa ao modelo de riscos proporcionais de Cox, no entanto e como veremos na análise de resultados este modelo apresenta limitações quando nos encontramos perante amostras “demasiado” grandes.

## 3.2. Resultados

### 3.2.1. Aplicação do Modelo de Cox e Modelo de Cox estratificado

Para a aplicação de análise de sobrevivência neste estudo foi utilizada uma “mega-amostra” correspondente às mães que tiveram os seus filhos em Portugal, no período entre 1995 e 2009, com a dimensão total  $n = 1184762$ . É importante referir que para a obtenção de um modelo final em análise de sobrevivência é necessário percorrer um determinado caminho de forma a obter um modelo explicativo dos dados o mais parcimonioso possível. Desta forma, devemos começar por avaliar a significância de cada uma das variáveis individualmente, e de seguida incluir no modelo apenas aquelas que apresentem uma significância aceitável (usualmente tem-se por base um  $\alpha = 0,20$ ).

Construído o primeiro modelo deverá retirar-se do modelo uma a uma as variáveis não significativas a 0,05 (ou mesmo 0,10), e posteriormente voltar a incluir uma a uma as variáveis que tínhamos excluído anteriormente, no primeiro passo. Finalmente deverão ser incluídas as interações significativas. É, no entanto, importante destacar que nenhuma destas etapas é aceitável se o modelo não permitir validar os pressupostos dos riscos proporcionais, sendo também importante a verificação de *outliers* ou pontos influentes, e por fim a verificação da bondade de ajustamento do modelo.

Deste modo e seguindo o acima explicado, a análise por nós desenvolvida seguiu cada um dos passos referidos.

#### 3.2.1.1. Espaçamento entre o casamento e o nascimento do primeiro filho

No contexto da evolução demográfica das últimas décadas é importante compreender a evolução entre o espaçamento decorrido entre o *evento* casamento e o *evento* nascimento do primeiro filho. Assim, e com o objectivo de avaliar a evolução do espaçamento decorrido entre o casamento e o nascimento do primeiro filho, os dados

aqui analisados são dados truncados<sup>4</sup>, já que apenas se pretende analisar a evolução no caso em que existem casamento e as mulheres tiveram o seu primeiro filho. Para compreender a evolução no espaçamento entre o casamento e o nascimento do primeiro, as variáveis utilizadas foram:

- Instrução da Mãe,
- Condição perante o trabalho da Mãe,
- Categoria profissional da Mãe,
- Situação profissional,
- Grupo de idade da Mãe,
- Década do casamento,
- Ano do nascimento (apenas numa análise global).

Ainda de referir e como referido anteriormente que foram elaboradas recodificações de algumas variáveis, como o ano do nascimento e a idade da mãe, recodificados de acordo com os quartis (tabela 4). Já o ano do casamento, foi recodificado em décadas.

**Tabela 4:** Recodificação de Variáveis

	Quartis				
	0%	25%	50%	75%	100%
Ano do nascimento	1995	1997	2001	2004	2007
Idade da Mãe	16	25	27	30	54

Recodificadas todas as variáveis, e como referido anteriormente, deverá então iniciar-se a validação individual das variáveis para construção dos modelos. Na tabela 5, onde temos os valores  $p$ , podemos observar que quer num modelo global ou no caso dos modelos de Cox estratificado anualmente entre 1995 e 2007, as variáveis em análise apresentam valores altamente significativos, assumindo desta forma que todas as variáveis apresentam significância para a construção dos modelos.

<sup>4</sup> Para a análise do espaçamento entre o casamento e o nascimento do primeiro filho, foram excluídos da análise os anos de 1999, 2008 e 2009 já que a informação disponibilizada pelo Instituto Nacional de Estatística apresentava alguns erros e ser necessária a sua correção.

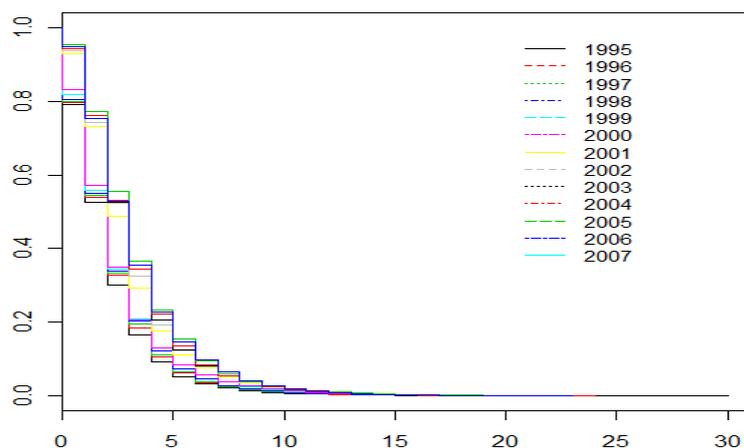
**Tabela 5:** Valores  $p$  para avaliar quais as variáveis significativas a incluir nos modelos anuais

	Ano Nascimento <sup>5</sup>	Nível de Instrução da Mãe	Condição perante o trabalho da Mãe	Profissão da Mãe	Situação Profissional da Mãe	Grupo de Idade da Mãe	Década do Casamento
Global	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
1995	-	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
1996	-	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
1997	-	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
1998	-	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
2000	-	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
2001	-	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
2002	-	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
2003	-	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
2004	-	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
2005	-	0.050	<0.001	0,040	<0.001	<0.001	<0.001
2006	-	0.020	<0.001	0,020	<0.001	<0.001	<0.001
2007	-	<0.001	<0.001	0,110	0,040	<0.001	<0.001

Ainda antes da obtenção de um modelo explicativo, é interessante verificar se os resultados da aplicação da análise de sobrevivência ao comportamento de fecundidade em Portugal, correspondem à informação empírica apresentada anteriormente na revisão da literatura, bem como na análise descritiva da situação de fecundidade em Portugal. Desta forma, seguem-se primeiramente as representações gráficas das variáveis em estudo, para o modelo global (figura 11) realizado através do modelo de Cox. Na figura 11, ao observarmos o ano de nascimento para o período em análise, verificamos que os comportamentos apresentam probabilidade de sobrevivência que evoluem para idade mais elevadas, num movimento das curvas para a direita. Desta forma e no contexto de um modelo global, podemos verificar que a probabilidade do evento de interesse, ou seja, o casamento e o nascimento do primeiro filho têm ocorrido em idades superiores deslocando assim a análise do seu espaçamento para estas idades. Se no início do período (1995) em análise a probabilidade de ocorrer o evento – nascimento do primeiro filho após o casamento – centrava-se ao longo dos primeiros anos de casamento além dos 5 anos, no final do período, em 2007 a probabilidade de ter o primeiro filho centrava-se de facto até aos 5 anos após o matrimónio.

<sup>5</sup> Covariável utilizada somente no modelo global.

**Figura 11:** Ano de nascimento do primeiro filho após o casamento no período entre 1995 e 2007



Por outro lado, é também importante, de uma forma geral, analisar o comportamento das restantes covariáveis. Na figura 12 encontramos a representação gráfica do estimador de Kaplan-Meier para a instrução da mãe, onde podemos observar que apesar de não se verificarem muitas diferenças entre as curvas de sobrevivência apresentadas na figura, podemos verificar que são as mães "sem instrução" que esperam menos tempo até ao nascimento do 1º filho, seguidas das que têm o ensino básico, secundário e superior. Assim, e como referido anteriormente na revisão da literatura, são as mães com nível de educação mais elevado que esperam mais tempo até ao nascimento do 1º filho.

Relativamente à condição perante o trabalho das mães (figura 13), verifica-se que as mães que se encontram em situação de desemprego, após o casamento, esperam menos tempo para ter o seu primeiro filho. Todavia, são as mães empregadas que estendem mais o tempo decorrido até ao nascimento do seu primeiro filho.

**Figuras 12 e 13:** Representação gráfica da instrução da mãe e a sua condição perante o trabalho no período entre 1995 e 2007.

Figura 12: Instrução da mãe

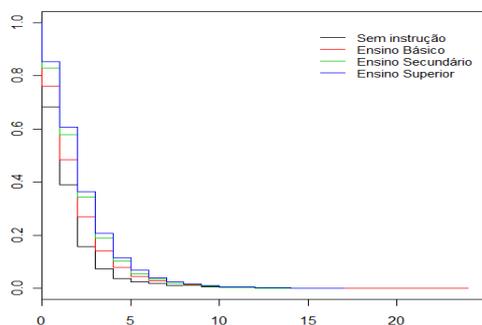
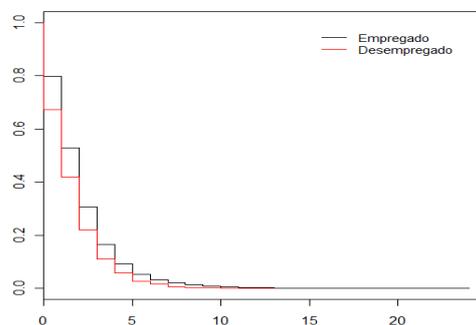


Figura 13: Condição perante o trabalho



Ao observarmos a distribuição gráfica da categoria profissional da mãe (figura 14), verificamos que são as mães da categoria profissional 3<sup>6</sup> que esperam menos tempo até ao nascimento do seu primeiro filho, encontrando-se em situação oposta as profissionais da categoria 1. Relativamente à situação profissional das mães (figura 15), observa-se um comportamento muito semelhante entre as mães empregadoras e as mães trabalhadoras por conta própria que adiam por mais tempo o nascimento do primeiro filho após o seu casamento.

**Figuras 14 a 15:** Representação gráfica da distribuição da categoria e da situação profissional da mãe no período entre 1995 e 2007.

Figura 14: Categoria Profissional

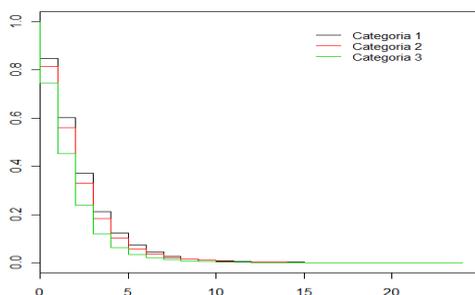
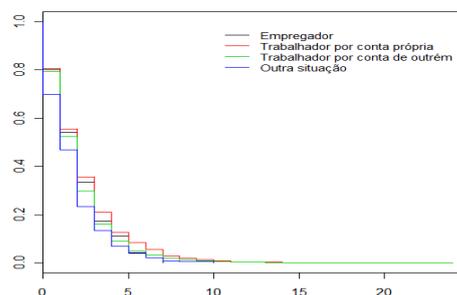


Figura 15: Situação Profissional



<sup>6</sup> Como referido no capítulo 2, à categoria profissional 3 correspondem Agricultores e trabalhadores qualificados da agricultura e pescas; Operários, artífices e trabalhadores similares; Operadores de instalações e máquinas e trabalhadores da montagem; Trabalhadores não qualificados. Na categoria profissional 2 enquadram-se, Técnicos e profissionais de nível intermédio; Pessoal administrativo e similares; Pessoal dos serviços e vendedores. Já na categoria profissional 1 encontram-se, Membros das forças armadas; Quadros superiores da administração pública, dirigentes e quadros superiores de empresas; Especialistas das profissões intelectuais e científicas.

Por fim, e ainda relativamente ao modelo global temos a representação gráfica do grupo de idade das mães, bem como a década relativa ao casamento, no período em análise. Observa-se na figura 16 que a probabilidade de as mães terem o seu filho depois dos 5 anos de casamento é apenas significativas, em mães com idade superiores a 30 anos. Ainda de destacar que as mães no grupo de idade 28-30 anos têm uma probabilidade superior às mães entre os 31 e 54 anos de nos primeiros um a dois anos terem o seu primeiro filho. Já no que diz respeito à década do casamento, observa-se de certa forma o que acabamos de referir na análise da idade mãe, isto é, a probabilidade de o evento ocorrer é superior para as mães que realizaram o seu casamento na década de 70 (as mães mais velhas), é possível verificar que as mães que realizaram os seus casamentos na década de 70 esperam até aproximadamente 20 anos após o seu casamento para o nascimento do seu primeiro filho. Por outro lado, as mães casadas na década de 80 concentram os nascimentos até aos oito anos após o casamento. E as mães casadas nas duas décadas seguintes (90 e 00) concentram o nascimento do seu primeiro filho nos anos imediatamente após o casamento. De referir ainda que a distribuição gráfica do comportamento de acordo com as décadas 70 ( $n = 104$ ), 80 ( $n = 5731$ ), 90 ( $n = 230442$ ) e 00 ( $n = 162784$ ) poderá estar condicionada com o número de eventos referentes a cada uma das décadas em causa. No entanto, a distribuição dos nascimentos de acordo com as décadas apresentadas é importante tendo em conta as evoluções decorridas neste período na fecundidade em Portugal.

**Figuras 16 a 17:** Representação gráfica da distribuição da idade da mãe e décadas do casamento no período entre 1995 e 2007.

Figura 16: Grupo de Idade da Mãe

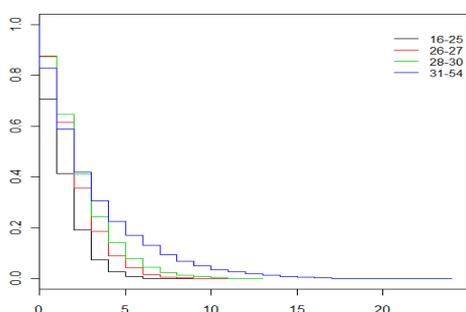
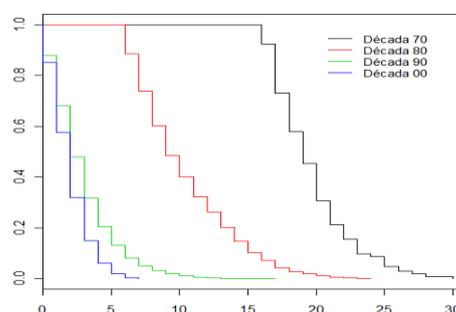


Figura 17: Ano do casamento em Décadas



Após uma apresentação do comportamento das variáveis explicativas da evolução da fecundidade em Portugal durante o período 1995 a 2007, é importante iniciar uma abordagem de modelo de Cox estratificado. Apresentamos, assim, de seguida a evolução quinquenal para cada uma das covariáveis - nível de instrução, condição perante o trabalho, categoria profissional, situação profissional, grupo de idade da mãe e ano do casamento em décadas.

Ao observar a evolução do impacto da instrução da mãe nos anos 1995, 2000 e 2005 (figura 18 a, b, c) verificamos que não existem diferenças significativas entre os anos, sendo que a maioria das mães tem o seu primeiro filho nos primeiros 5 anos de casamento. No entanto é de destacar uma vez mais que são as mulheres com mais instrução que “adiam” por mais tempo a probabilidade de ter o seu primeiro filho. De destacar ainda que no ano de 2005 as mulheres sem escolaridade apresentam uma situação diferente dos anos anteriores, sendo estas a apresentar um maior adiamento da fecundidade; no entanto, esta alteração pode também estar relacionada com o aumento dos níveis de escolaridade, diminuindo a amostra de mães sem instrução nos últimos anos.

**Figura 18:** Representação gráfica do nível de instrução da mãe nos anos de 1995, 2000 e 2005.

Figura 18a: Nível de instrução no ano 1995

Figura 18b: Nível de instrução no ano 2000

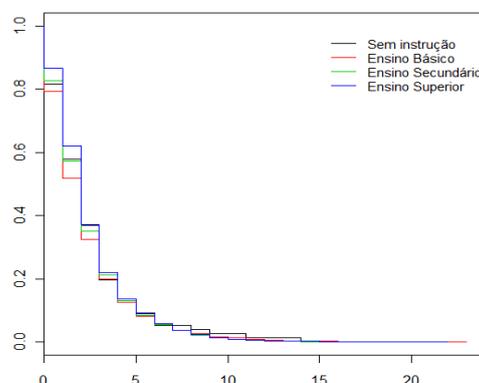
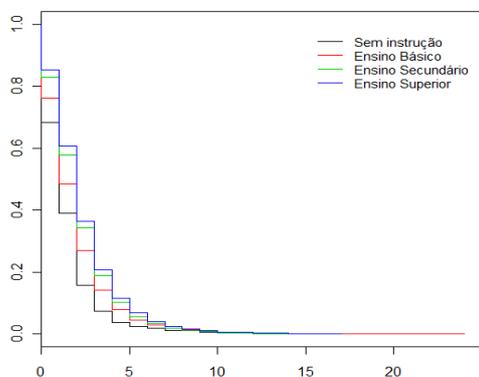
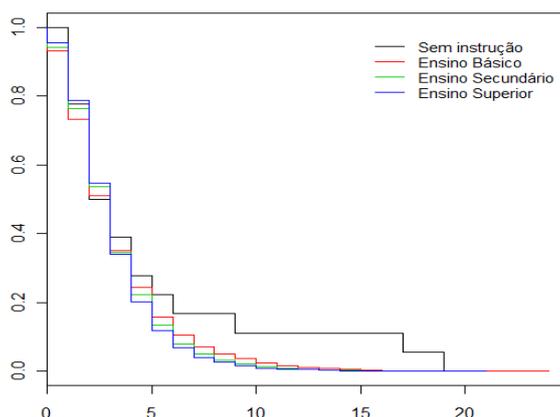


Figura 18c: Nível de instrução no ano 2005



Relativamente à condição perante o trabalho das mães (figura 19), podemos observar a não existência de alterações nos comportamentos, no entanto devemos destacar que no ano de 2005 a probabilidade de as mães empregadas ou desempregadas terem o seu primeiro filho imediatamente a seguir ao seu casamento aumentou em comparação com os dois anos anteriores. Por outro lado, e embora o adiamento seja mais pronunciado no caso de mulheres empregadas, este adiamento tem aumentado e tem também aumentado para as mulheres desempregadas.

**Figura 19:** Representação gráfica da condição perante o trabalho da mãe nos anos de 1995, 2000 e 2005.

Figura 19a: Condição perante o trabalho no ano 1995

Figura 19b: Condição perante o trabalho no ano 2000

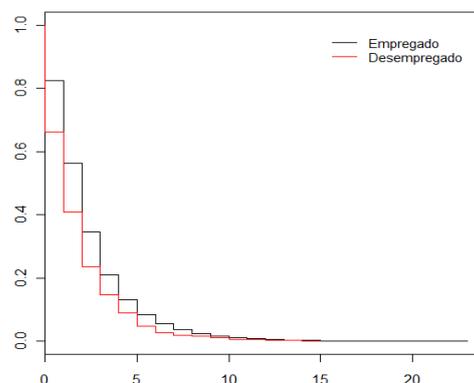
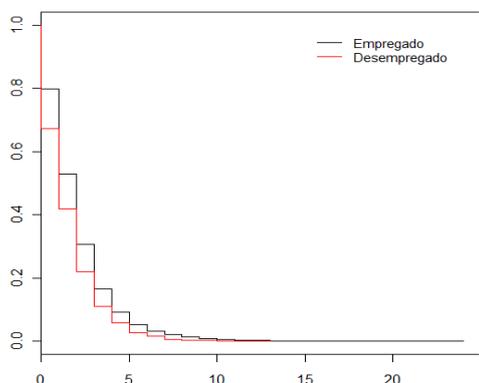
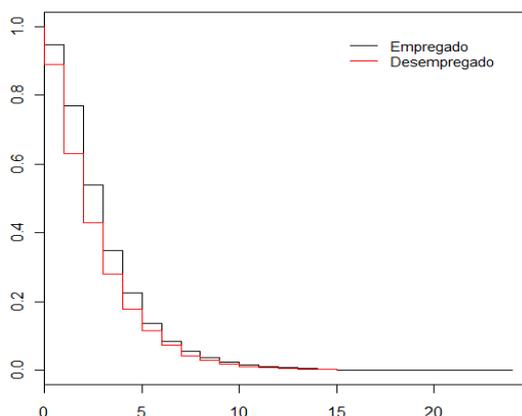


Figura 19c: Condição perante o trabalho no ano 2005



Já no que diz respeito à categoria profissional da mãe (figura 20) verificamos um padrão de nascimentos ocorridos nos primeiros anos, e por outro lado, podemos também observar, que a probabilidade de o evento ocorrer num menor espaço de tempo aumenta no ano de 2005. Como referido anteriormente aquando da análise do modelo global podemos também observar que as mães pertencentes à categoria profissional 1 (aquela que engloba mães com níveis de educação mais elevados) apresentam probabilidades de o evento de interesse ocorrer mais elevadas revelando que este *momentum* está a ser adiado.

**Figura 20:** Representação gráfica da categoria Profissional da mãe nos anos de 1995, 2000 e 2005.

Figura 20a: Categoria Profissional da mãe em 1995

Figura 20b: Categoria Profissional da mãe em 2000

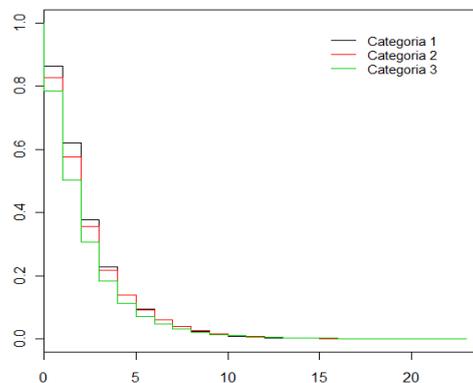
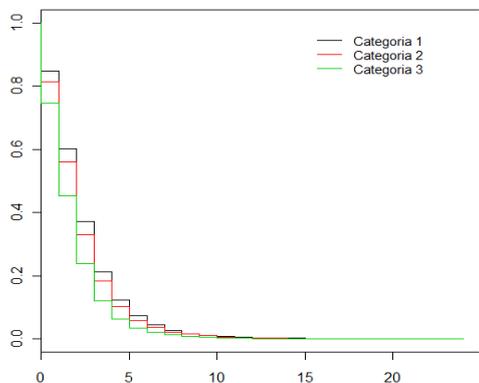
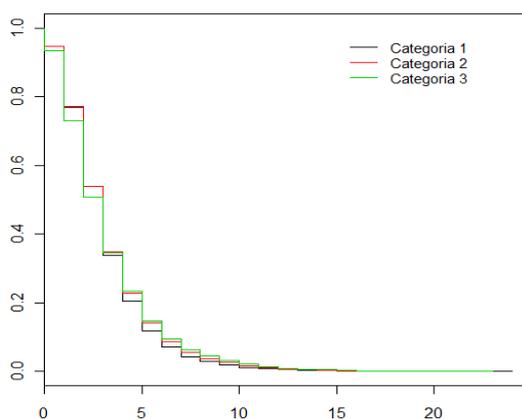


Figura 20c: Categoria Profissional da mãe em 2005



Relativamente à situação profissional da mãe (figura 21), e à semelhança do que acontecia no modelo global que incluía o comportamento de todos os anos em análise, observa-se um comportamento semelhante entre as mães empregadoras e as mães trabalhadoras por conta própria que adiam por mais tempo o nascimento do primeiro filho após o seu casamento.

**Figura 21:** Representação gráfica da situação Profissional da mãe nos anos de 1995, 2000 e 2005.

Figura 21a: Situação profissional das mães em 1995

Figura 21b: Situação profissional das mães em 2000

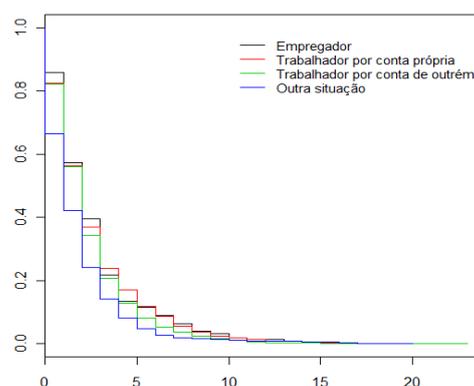
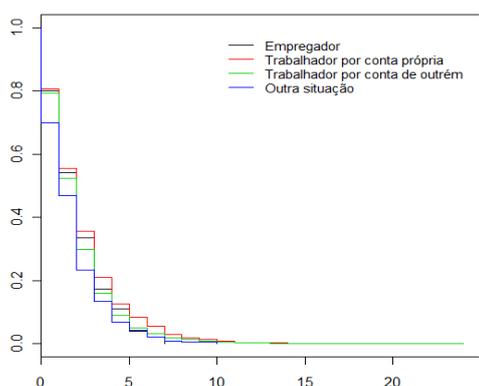
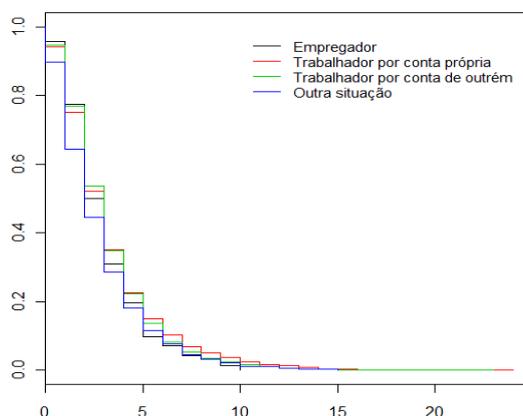


Figura 21c: Situação profissional das mães em 2005



Na figura 22 podemos observar o comportamento do grupo de idade da mãe nos anos de 1995, 2000 e 2005. Verifica-se que nestes quinquênios os comportamentos relativamente à idade não sofreram alterações de maior, sendo, no entanto, de destacar uma vez mais que no ano de 2005 as probabilidades do evento acontecer no primeiro ano a seguir ao casamento são mais elevadas que nos anos anteriores, revelando uma mudança de comportamento, ainda que ligeira. É ainda importante referir que os comportamentos nos três anos em análise não são muito diferentes até aos 5 anos após o

casamento; porém, após este período as probabilidades diferenciam-se de ano para ano, sendo que no ano de 2005 a probabilidade de uma mulher ter o seu primeiro filhos após o casamento no grupo de idades 31-54 é superior à verificada nos anos anteriores.

**Figura 22:** Representação gráfica do grupo de Idade da mãe nos anos de 1995, 2000 e 2005.

Figura 22a: Grupo de idade da mãe em 1995

Figura 22b: Grupo de idade da mãe em 2000

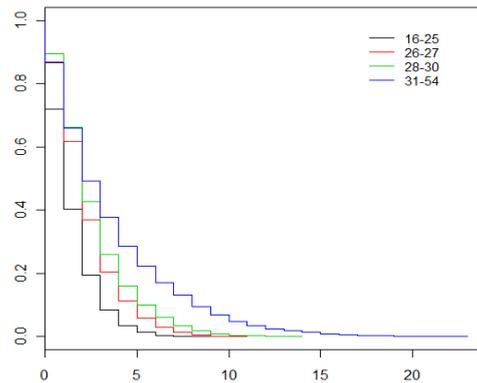
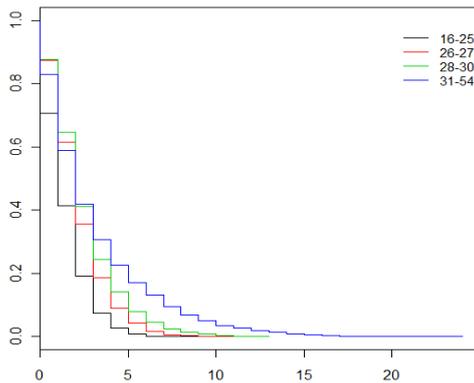
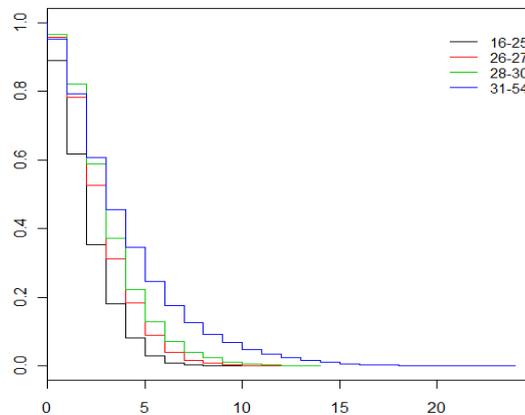


Figura 22c: Grupo de idade da mãe em 2005



Com base nos valores  $p$  apresentados anteriormente (tabela 6), onde com significância de  $\alpha = 0,20$ , se validaram todas as variáveis individualmente, considerando que todas as variáveis são significativas para a construção de modelos. Procedeu-se assim à fase de modelação, onde se construíram modelos explicativos da evolução do comportamento da fecundidade entre o casamento e o nascimento do primeiro filho.

Assim, na tabela 7, encontra-se a síntese dos modelos finais de acordo com cada ano em análise com os seus respectivos Hazard Rates (HR ou  $exp(\beta)$ ) e onde se destacam diferentes comportamentos. Os modelos finais para cada ano em análise apresentam semelhanças nas covariáveis explicativas do comportamento de fecundidade. No entanto, é importante destacar que o modelo correspondente ao ano de 1995 é o único em que todas as variáveis são significativas. Por outro lado, é também de salientar que apenas nos anos de 1996/7/8 e 2003/4/5 se apresentam modelos com interações entre as covariáveis. É também importante distinguir que a covariável idade da mãe é comum em todos os modelos finais apresentados, revelando a importância desta covariável no comportamento da fecundidade em Portugal.

Contudo, é importante analisar os valores obtidos. Assim, e tendo primeiramente o ano de 1995 como exemplo, na covariável instrução da mãe, verifica-se que relativamente à categoria profissional 1, em comparação com a categoria de referência, esta apresenta uma probabilidade de -20% de ocorrer o evento, ou seja, no contexto deste estudo significa que para as mulheres o espaçamento entre o casamento e o nascimento do primeiro filho é inferior às da mulher da categoria de referência. A situação não difere quando se observam as restantes categorias desta covariável.

Por outro lado, quando observamos a covariável condição perante o trabalho (empregada/desempregada) ao longo do período, verificamos que o seu comportamento até ao ano de 2007 foi crescente; se em 1995 existia mais 10% de probabilidade de ocorrer o evento (ter o primeiro filho) no caso das mães desempregadas, e em 2006, 14% de probabilidade, já em 2007 estes valores invertem-se e a situação altera-se sendo que passou a existir uma probabilidade de -15% de que as mães empregadas tenham os seus filhos mais cedo.

Ainda neste contexto ao observarmos a idade da mãe no ano de 1995, verificamos que o aumento de idade significa a diminuição da probabilidade de ter filhos (-31% para o primeiro grupo de idade, -49% para o segundo grupo de idades e -49% no último grupo de idade da mãe).

Neste contexto é ainda importante destacar que nestes modelos finais explicativos da evolução do espaçamento entre o casamento e o nascimento do primeiro filho, verificaram-se apenas duas interações entre as covariáveis, a primeira entre a condição perante o trabalho e o grupo de idade da mãe nos anos de 1996, 1997 e 1998, e a

segunda interação entre a categoria profissional e a idade da mãe nos anos de 2003, 2004 e 2005.

**Tabela 6:** Valores de HR dos modelos finais de por cada ano em análise

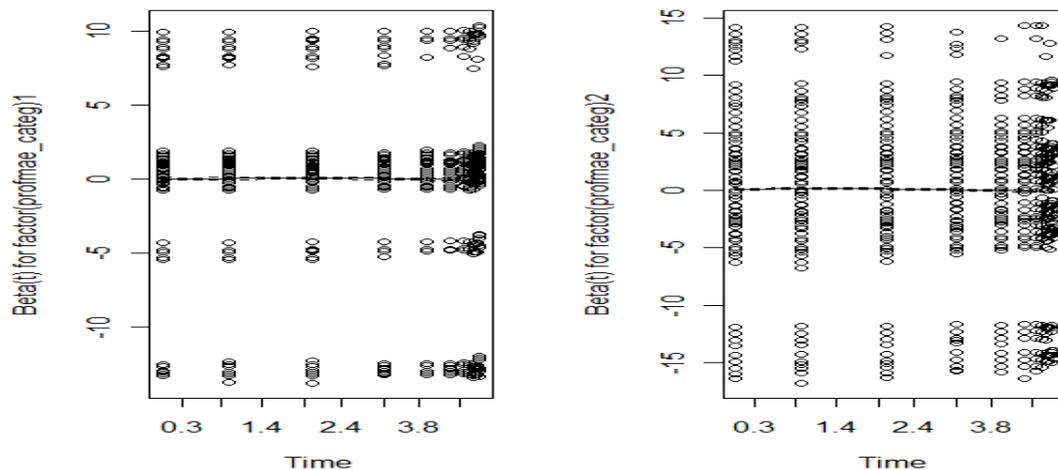
	1995	1996	1997	1998	2000	2001	2002	200	2004	2005	2006	2007
isnrmae_categ1	0,80	-	-	-	-	-	-	-	-	-	-	-
isnrmae_categ2	0,83	-	-	-	-	-	-	-	-	-	-	-
isnrmae_categ3	0,93											
cdtbmae	1,10	1,18	1,16	1,84	1,18	1,14	1,16	1,11	-	-	1,14	0,85
profmae_categ1	1,03	0,90	-	-	0,86	0,86	-	0,94	0,88	0,86	0,86	0,75
profmae_categ2	1,08	0,92	-	-	0,82	0,81	-	0,9	0,86	0,83	0,77	1,09
idademae_categ1	0,61	0,62	0,65	0,64	0,62	0,62	0,67	0,79	0,79	0,72	0,7	0,73
idademae_categ2	0,51	0,50	0,57	0,55	0,51	0,51	0,58	0,62	0,65	0,60	0,59	0,61
idademae_categ3	0,41	0,4	0,45	0,44	0,37	0,37	0,45	0,44	0,48	0,47	0,45	0,45
cdtbmae3:idademae_categ1	-	0,92	0,93	0,85	-	-	-	-	-	-	-	-
cdtbmae3:idademae_categ2	-	0,88	0,89	0,92	-	-	-	-	-	-	-	-
cdtbmae3:idademae_categ3	-	0,77	0,8	0,82	-	-	-	-	-	-	-	-
profmae_categ1:idademae_categ1	-	-	-	-	-	-	-	0,87	0,92	1,00	-	-
profmae_categ2:idademae_categ1	-	-	-	-	-	-	-	0,79	0,81	0,83	-	-
profmae_categ1:idademae_categ2	-	-	-	-	-	-	-	0,9	0,98	1,03	-	-
profmae_categ2:idademae_categ2	-	-	-	-	-	-	-	0,83	0,84	0,93	-	-
profmae_categ1:idademae_categ3	-	-	-	-	-	-	-	0,92	0,98	0,99	-	-
profmae_categ2:idademae_categ3	-	-	-	-	-	-	-	0,87	0,92	0,89	-	-

Os resultados obtidos nestes modelos explicativos e parcimoniosos no que diz respeito a análise do espaçamento entre o casamento e o nascimento do primeiro filho, do ponto de vista do contexto demográfico da evolução da fecundidade em Portugal apresentam-se significativos e vão de encontro aos resultados empíricos. Contudo, e não excluindo de modo algum a abordagem estatística que caracteriza este trabalho, é importante verificar se estes modelos apresentados satisfazem os pressupostos estatísticos associados à análise de sobrevivência e nomeadamente aos modelos proporcionais de Cox.

Assim, apresentamos de seguida a representação gráfica associado aos pressupostos do modelo de Cox, e tendo em conta o número de covariáveis (7) em análise bem como o número de anos em análise (12 anos) apresentaremos apenas os resultados para um dos anos (1995).

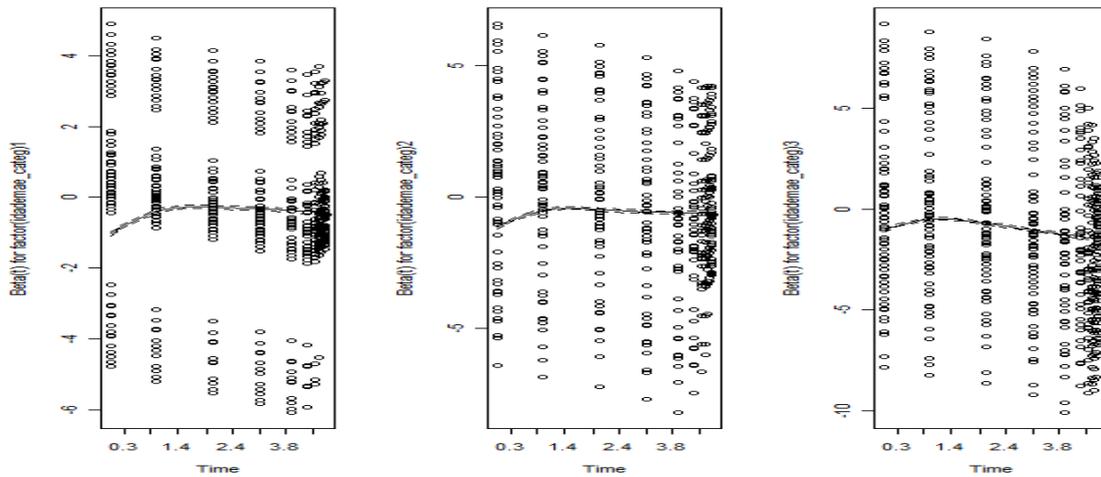
Deste modo, e no que diz respeito aos resíduos de *schoenfeld* padronizados, apresentamos os resultados para as covariáveis profissão da mãe e idade da mãe. Na figura 23 podemos observar o comportamento para resíduos de Schoenfeld padronizados da covariável categoria profissional da mãe, verificando que a distribuição não é de forma alguma aleatória, verificando-se uma tendência na distribuição dos resultados.

**Figura 23:** Resíduos de Schoenfeld padronizados para a covariável categoria profissional da mãe no ano de 1995



Relativamente à representação dos resíduos de *schoenfeld* padronizados para a covariável grupo de idade das mães (figura 24) verifica-se a presença de um padrão mais aleatório dos dados; no entanto, o resultado desta representação gráfica não é satisfatório da proporcionalidade das covariáveis.

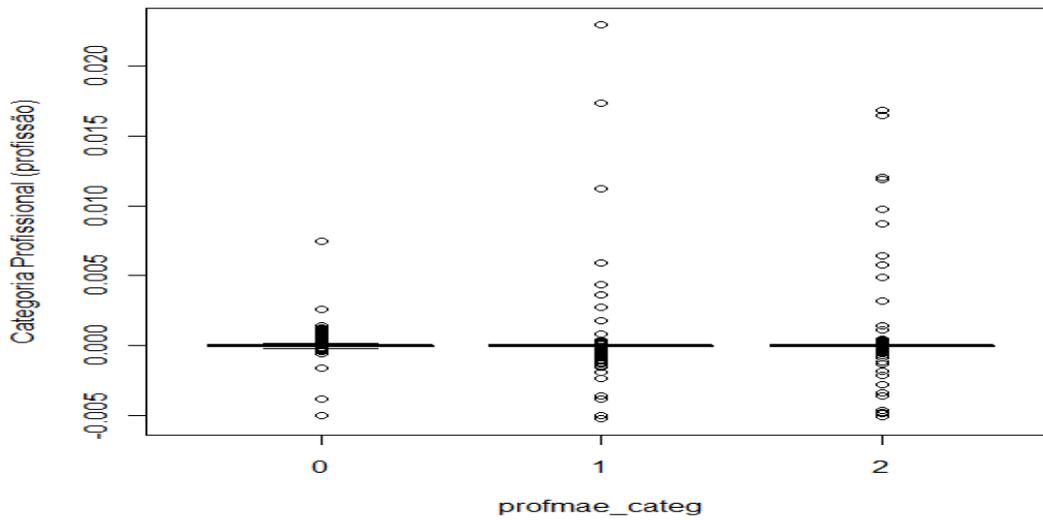
**Figura 24:** Resíduos de Schoenfeld padronizados para a covariável grupo de idade da mãe no ano de 1995



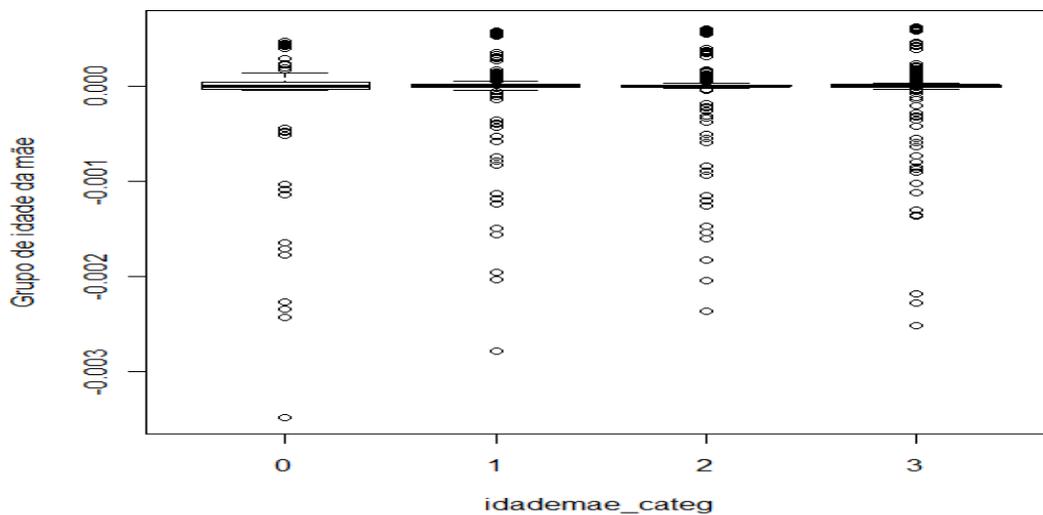
A não verificação dos riscos proporcionais nas covariáveis apresentadas através dos resíduos de *schoenfeld* padronizados, verifica-se também nas restantes covariáveis, levando-nos a crer que no futuro diferentes abordagens se devem ter em conta, como, por exemplo, a criação de outro tipo de modelo (modelo de tempo de vida acelerado), ou até uma melhor combinação entre as técnicas demográficas e estatísticas.

Contudo, é também importante a observação dos resíduos *score*, que permitem a verificação da influência de cada observação no ajuste do modelo. Desta forma, e novamente para o ano de 1995 e covariáveis profissão da mãe e grupo de idade da mãe apresentamos os resultados gráficos nas figuras 25 e 26. Nestas representações gráficas dos resíduos *score* verifica-se a existência de inúmeras observações influentes que poderão influenciar os resultados finais obtidos. No entanto, e considerando que se este trabalho é representativo de uma população e não de uma amostra, não fará sentido a remoção destas observações consideradas influentes ou *outliers*.

**Figura 25:** Resíduos Score para a covariável categoria profissional da mãe no ano de 1995



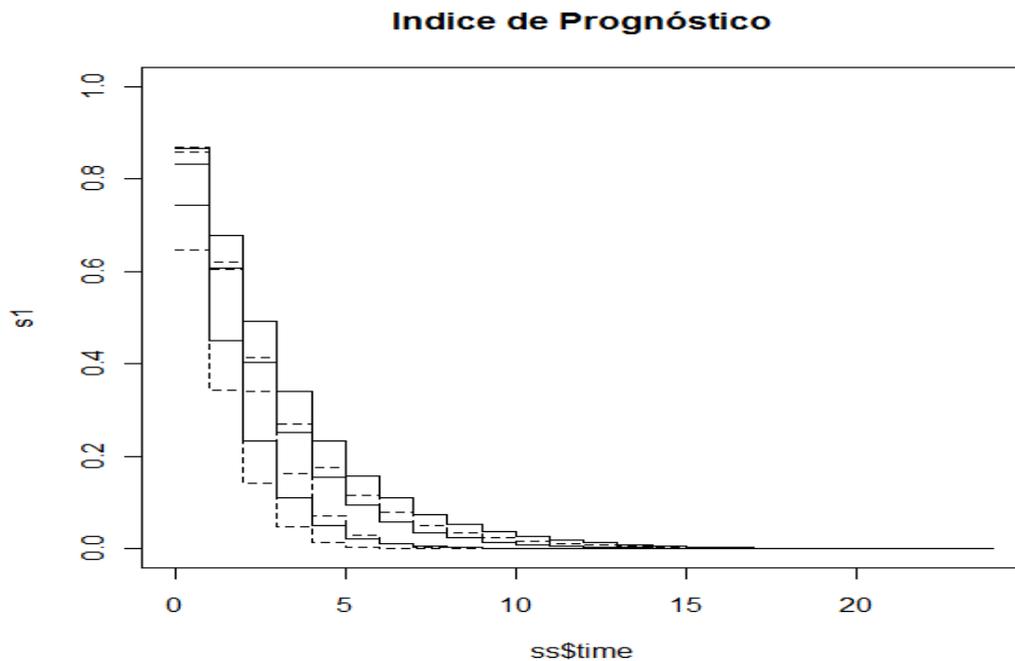
**Figura 26:** Resíduos Score para a covariável grupo de idade da mãe no ano de 1995



Embora as covariáveis utilizadas para explicar a evolução no espaçamento entre o casamento e o nascimento do primeiro filho, apresentem resultados não satisfatórios para se considerarem modelos de riscos proporcionais de Cox, é ainda necessário realizar um teste de índice de prognóstico de forma a verificar a adequabilidade dos modelos finais. A representação gráfica do índice de prognóstico (figura 27) possibilita a observação nas linhas sólidas o modelo final ajustado e nas linhas tracejadas as

estimativas Kaplan-Meier. Assim e embora os modelos finais pareçam não assumir os riscos proporcionais, ao observar o índice de prognóstico verificamos que parece de facto existir um bom ajustamento dos modelos.

**Figura 27:** Índice de Prognóstico para o modelo de 1995



Como referido anteriormente, e de acordo com os resultados obtidos nos pressupostos associados ao modelo de riscos proporcionais de Cox, não foram obtidos resultados significativos para afirmar a proporcionalidade dos modelos. Contudo, o resultado obtido com o índice de prognóstico é revelador de que os modelos são de facto explicativos da evolução entre o casamento e o nascimento do primeiro filho.

Na evolução do espaçamento entre o casamento e o nascimento do primeiro filho, concluímos assim que no período de 1995 a 2007 as covariáveis que influenciam o espaçamento e a decisão de ter filhos são a condição perante o trabalho, a idade da mãe e, por fim, a categoria profissional. Sendo que na condição perante o trabalho se observou que as mães desempregadas esperam menos tempo entre o casamento e o nascimento sendo a sua probabilidade para a ocorrência do evento inferior às mulheres em situação de empregadas. Relativamente à idade da mãe, observa-se que são as mulheres mais velhas que mais esperam até ao nascimento do primeiro filho após o

casamento. Por fim, e no que diz respeito à categoria profissional, observou-se que mães da categoria profissional 1 esperam mais tempo que as restantes sendo que são as mães da categoria profissional 3 as que menos tempo esperam.

### 3.2.1.2. Espaçamento entre o nascimento do primeiro e o segundo filho

A diminuição no número de filhos e o adiamento do nascimento dos mesmos tem sido questão fulcral na análise de fecundidade nas últimas décadas. Neste contexto, é necessário compreender o comportamento das mães portuguesas e o que influencia o espaçamento entre o nascimento do primeiro e do segundo filho. Na aplicação da análise de sobrevivência os dados referentes aos anos de 2006 e 2007 foram truncados por se apresentarem anómalos. Para compreender a evolução no espaçamento entre o nascimento do primeiro e do segundo filho, as variáveis utilizadas foram:

- Instrução da Mãe,
- Condição perante o trabalho da Mãe,
- Categoria profissional da Mãe,
- Situação profissional da Mãe,
- Grupo de idade da Mãe,
- Filiação,
- Ano do nascimento (apenas numa análise global).

À semelhança do caso anterior foram elaboradas recodificações de algumas covariáveis, o ano do nascimento foi agrupado (1995/99, 2000/01, 2002/03 e 2004/05) e a idade da mãe recodificada de acordo com os seus quantis (tabela 7).

**Tabela 7:** Recodificação de Variáveis

	0%	25%	50%	75%	100%
Idade da Mãe	13	25	28	32	54

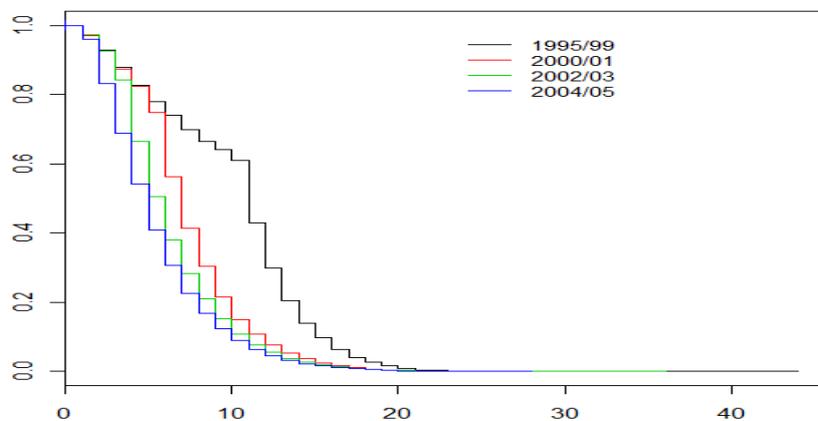
Uma vez mais a análise inicia-se pela validação individual das variáveis com um  $\alpha = 0,20$  (tabela 8). Contrariamente à análise anterior aqui observamos a presença de uma covariável que não se apresenta significativa em alguns dos anos em análise. A covariável “condição perante o trabalho” apresenta-se como não significativa nos anos de 2000 a 2003 e em 2005. Este resultado parece indicar que, nestes anos referenciados, o facto de as mães estarem ou não desempregadas não influenciou a sua decisão para terem o segundo filho. Iremos, no entanto, mais à frente verificar se de facto esta covariável deixa de ser significativa.

**Tabela 8:** Valores  $p$  para avaliar variáveis significativas a incluir nos modelos anuais

	Ano do Nascimento	Filiação	Nível de Instrução da Mãe	Condição perante o trabalho da Mãe	Profissão da Mãe	Situação Profissional da Mãe	Grupo de Idade da Mãe
Global	<0.001	<0.001	<0.001	0,040	<0.001	<0.001	<0.001
1995	-	<0.001	<0.001	0,030	<0.001	<0.001	<0.001
1996	-	<0.001	<0.001	0,020	<0.001	<0.001	<0.001
1999	-	<0.001	<0.001	0,110	<0.001	<0.001	<0.001
2000	-	<0.001	<0.001	0,940	<0.001	<0.001	<0.001
2001	-	<0.001	<0.001	0,640	<0.001	<0.001	<0.001
2002	-	<0.001	<0.001	0,940	<0.001	<0.001	<0.001
2003	-	<0.001	<0.001	0,820	<0.001	<0.001	<0.001
2004	-	<0.001	<0.001	0,010	<0.001	<0.001	<0.001
2005	-	<0.001	<0.001	0,900	<0.001	0,050	<0.001

Tal como na análise anterior sobre o espaçamento entre o casamento e o primeiro filho, antes da obtenção de um modelo explicativo, é interessante observar se os resultados da aplicação da análise de sobrevivência ao comportamento de fecundidade em Portugal, na análise do espaçamento entre o nascimento do primeiro e segundo filho, correspondem à informação empírica apresentada anteriormente na revisão da literatura. Na figura 28, podemos observar a evolução dos comportamentos no espaçamento entre o nascimento do primeiro e segundo filho, verificando que o período entre 1995 e 1999 é aquele que apresenta as probabilidades maiores após os 10 anos decorridos do nascimento do primeiro filho. Este comportamento revela uma alteração dos comportamentos descrito anteriormente, à medida que o número de nascimentos está a diminuir, diminui também o tempo decorrido entre nascimentos.

**Figura 28:** Distribuição dos nascimentos do segundo filho das mães portuguesas no período entre 1995 e 2005



Na figura 29, podemos observar o comportamento das mães no período em análise de acordo com o nível de instrução das mesmas. Assim, é de destacar que as probabilidades de ter um segundo filho são mais elevadas nos primeiros cinco anos após o nascimento do primeiro filho. Observa-se também que as mães com o ensino básico apresentam nos primeiros 10 anos, maiores probabilidades de ter um segundo filho comparativamente com as mães das restantes categorias. São as mães com um nível de instrução “ensino superior” que apresentam as probabilidades mais baixas, podendo este comportamento revelar que estas mães têm o seu segundo filho mais cedo ou, por outro lado, optam por famílias mais pequenas.

No que diz respeito à condição das mães perante o trabalho (figura 30), observa-se que, na generalidade, as probabilidades de mais tempo de espera até ter o segundo filho são maiores para as mães desempregadas, com exceção do período entre um a três anos após o nascimento do primeiro filho, em que as probabilidades são mais elevadas para as mães empregadas.

Figuras 29 e 30: Representação gráfica da instrução da mãe e a sua condição perante o trabalho no período em análise.

Figura 29: Instrução da mãe

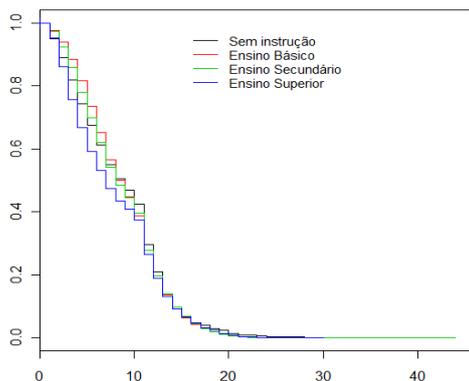
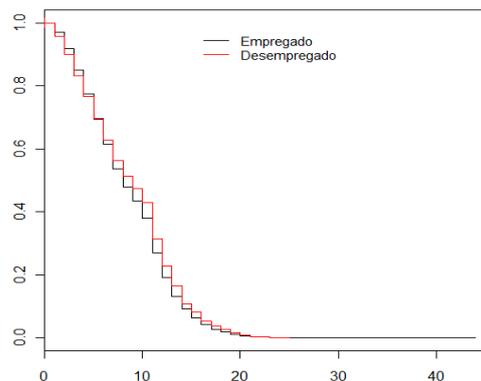


Figura 30: Condição perante o trabalho



Reflectindo o observado anteriormente relativamente ao “nível de instrução da mãe”, na análise da “categoria profissional” (figura 31) observa-se que as mães com profissões inseridas na categoria 3 apresentam probabilidades mais elevadas de terem o seu segundo filho, seguido das mães na categoria profissional 2 e por fim a categoria 1 que apresenta probabilidades mais baixas. Embora no primeiro ano após o nascimento do primeiro filho, a situação não seja muito diferente, depois deste período os comportamentos distinguem-se. No que concerne à “situação profissional” das mães (figura 32), embora o comportamento não seja diferente entre as diversas situações profissionais, principalmente nos primeiros anos após o nascimento do primeiro filho, é de salientar, no entanto, que as probabilidades de ter um segundo filho são mais elevadas para as mães trabalhadoras por conta de outrem.

**Figuras 31 e 32:** Representação gráfica da categoria profissional e situação profissional das mães no período em análise.

Figura 31: Categoria Profissional

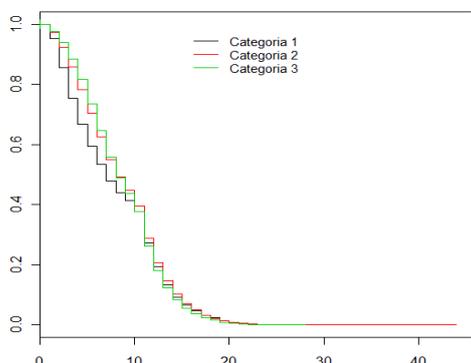
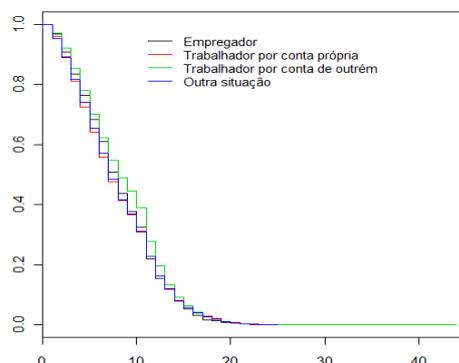


Figura 32: Situação Profissional



Ao observar a covariável “grupo de idade da mãe” (figura 33) para o período em análise, verificamos que existe, nos primeiros cinco anos, um comportamento muito semelhante entre as diferentes idades. Contudo, é o grupo de idades 13-25 que se destaca por ser aquele que após um período de 10 anos depois do nascimento do primeiro filho, apresenta probabilidades mais elevadas, no que diz respeito ao nascimento de um segundo filho. Este comportamento neste grupo de idades poderá reflectir a decisão de mães que ao nascimento do primeiro filho eram muito novas, e como tal decidem adiar até mais tarde o nascimento de um segundo filho. É igualmente de destacar que são as mães do grupo de idades 29-32 que apresentam probabilidades menores de ocorrência do evento.

Relativamente à “filiação dos nascimentos” (figura 34), verifica-se que no período em análise são as mulheres que têm os seus filhos fora do casamento que apresentam probabilidades mais elevadas de esperarem mais tempo até ao nascimento de um segundo filho, sendo também importante referir que as mulheres casadas prolongam por mais tempo esta decisão.

**Figuras 33 e 34:** Representação gráfica do grupo de idade da mãe e filiação no período em análise.

Figura 33: Grupo de Idade da Mãe

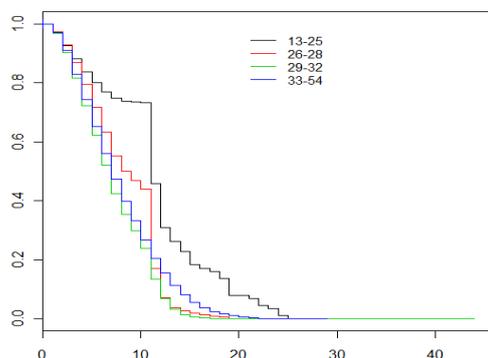
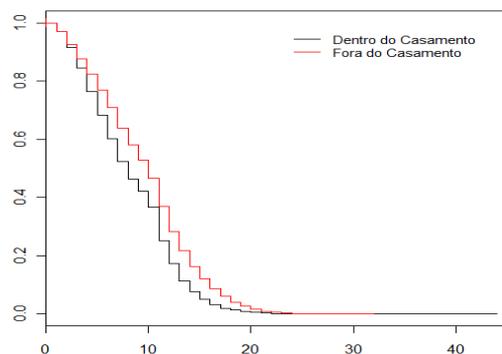


Figura 34: Filiação



Após uma apresentação do comportamento das covariáveis explicativas da evolução do espaçamento entre o nascimento do primeiro e do segundo filho, em Portugal, no período de 1995 a 2005, é importante iniciar uma abordagem de modelo de Cox estratificado. Como realizado anteriormente na análise do comportamento entre o casamento e o nascimento do primeiro filho, apresentamos de seguida a evolução quinquenal para cada uma das variáveis explicativas dos modelos - nível de instrução, condição perante o trabalho, categoria profissional, situação profissional, grupo de idade da mãe e filiação do nascimento.

Ao observar a evolução do comportamento através do nível de instrução da mãe (figura 35), verificamos que, no ano de 1995, as probabilidades de uma mãe ter o seu segundo filho eram significativamente altas até 10 anos após o nascimento do primeiro filho, e independentemente do nível de instrução. Todavia, destaca-se uma probabilidade mais elevada para as mães com o ensino básico e com o ensino secundário. Já no ano de 2000 estas probabilidades diminuem e cinco anos após o nascimento do primeiro filho, o comportamento é distintamente diferente do anterior, já que neste ano as mães diminuem as suas probabilidades de terem um segundo filho. No ano de 2005 é notório que as mães com o nível de ensino secundário são as que apresentam maior tempo de espera até ocorrer o evento de interesse, e sendo as mães com um nível de instrução mais elevado (ensino superior) as que apresentam probabilidades menores.

**Figura 35:** Representação gráfica do nível de instrução da mãe nos anos de 1995, 2000 e 2005.

Figura 35a: Nível de instrução no ano 1995

Figura 35b: Nível de instrução no ano 2000

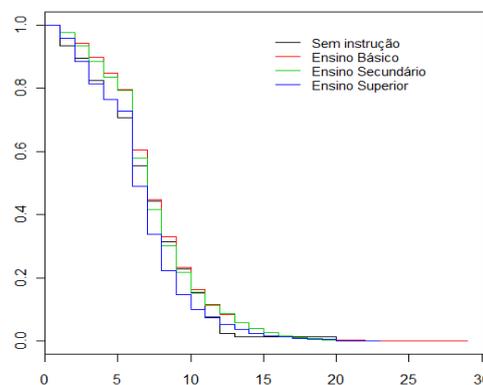
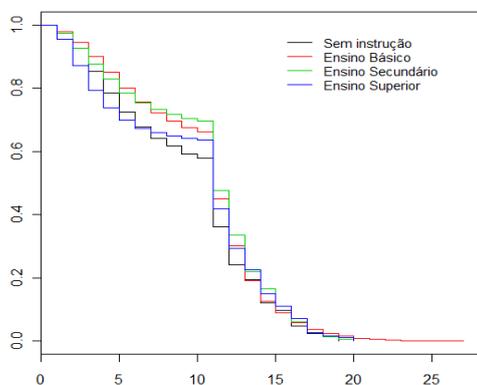
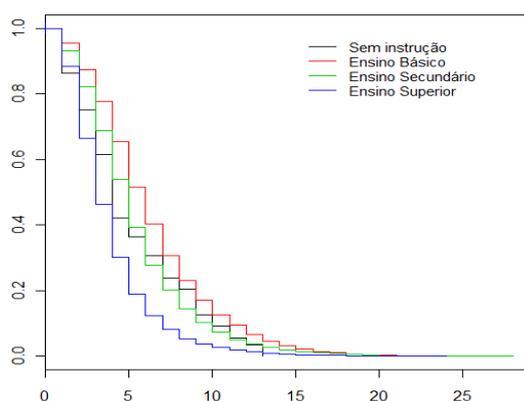


Figura 35c: Nível de instrução no ano 2005



Relativamente à “condição perante o trabalho da mãe” (figura 36), é de referir que não se verificam diferenças significativas entre as mães empregadas e as mães desempregadas. No entanto, o comportamento é muito semelhante no nível de instrução da mãe, observando-se no ano de 1995 um “adiamento” com elevadas probabilidades até aos 10 anos após o nascimento do primeiro filho, decrescendo para 5 anos no ano de 2000 e em 2005 verifica-se a diminuição da probabilidade de as mães terem o seu segundo filho. É ainda importante referir que para mães empregadas ou desempregadas o tempo de espera até à decisão de ter o segundo filho é praticamente igual.

**Figura 36:** Representação gráfica da condição perante o trabalho da mãe nos anos de 1995, 2000 e 2005.

Figura 36a : Condição perante o trabalho no ano 1995

Figura 36b: Condição perante o trabalho no ano 2000

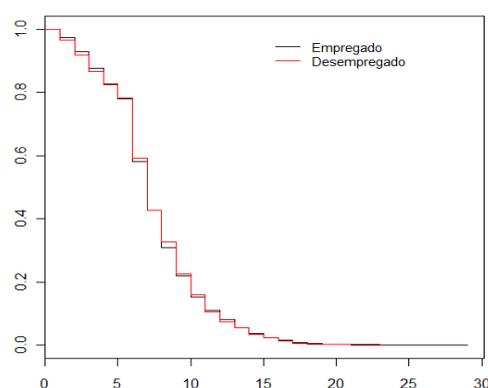
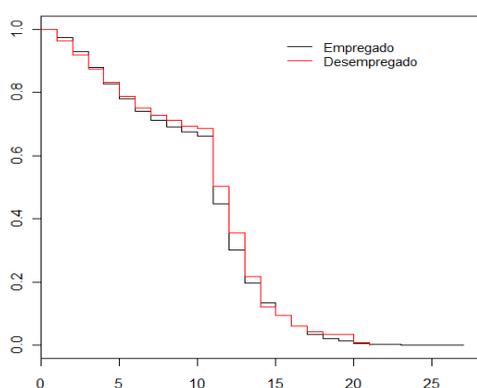
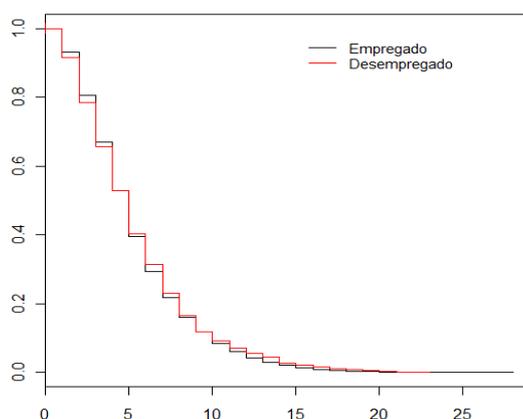


Figura 36c: Condição perante o trabalho no ano 2005



Ao observar a evolução do espaçamento entre o primeiro e o segundo filho de acordo com a “categoria profissional da mãe” (figura 37), podemos verificar que as mães pertencentes à categoria 1 são as únicas que se diferenciam das restantes, nomeadamente, em 1995 e 2000, visto que são as que apresentam uma probabilidade menor, reflectindo a probabilidade de mães com profissões melhor remuneradas esperarem menos tempo até ao nascimento de um segundo filho. Por outro lado, o comportamento das outras duas categorias não é distinto o suficiente para que nos permita de facto identificar diferenças de comportamento significativas. No entanto, no ano de 2000 verifica-se uma distanciação entre estas duas categorias e são as mulheres enquadradas na categoria profissional 3 que apresentam probabilidades maiores de esperarem mais tempo para o nascimento do segundo filho.

**Figura 37:** Representação gráfica da categoria profissional da mãe nos anos de 1995, 2000 e 2005.

Figura 37a: Categoria Profissional no ano 1995

Figura 37b: Categoria Profissional no ano 2000

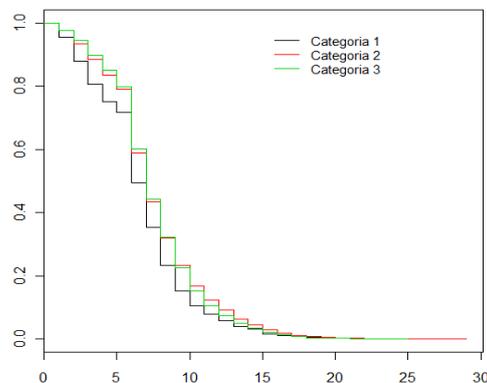
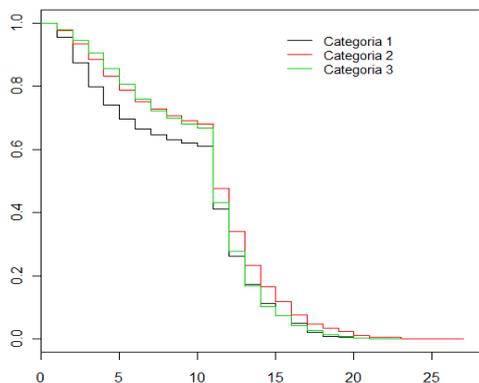
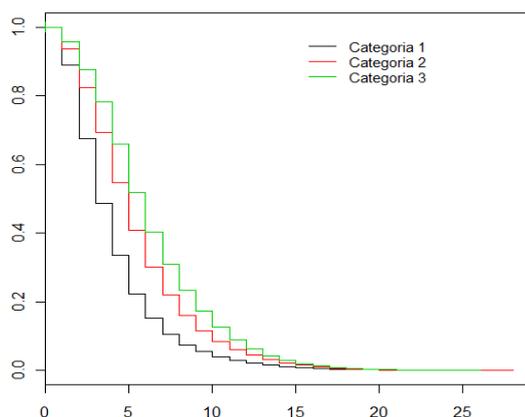


Figura 37c: Categoria Profissional no ano 2005



Também a “situação profissional das mães” (figura 38) apresenta a tendência referida anteriormente. O ano de 1995 apresenta um comportamento específico que não é visível nos restantes anos. Todavia, o comportamento é semelhante nos diferentes anos, isto é, não se verifica um comportamento distinto entre as várias categorias de situação profissional. De salientar também que a categoria que “ganha” maior destaque é a respeitante a “outra situação”, sendo que nesta categoria estão agregadas mulheres em situação de desemprego e domésticas, entre outras situações.

**Figura 38:** Representação gráfica da situação profissional da mãe nos anos de 1995, 2000 e 2005.

Figura 38a: Situação Profissional no ano 1995

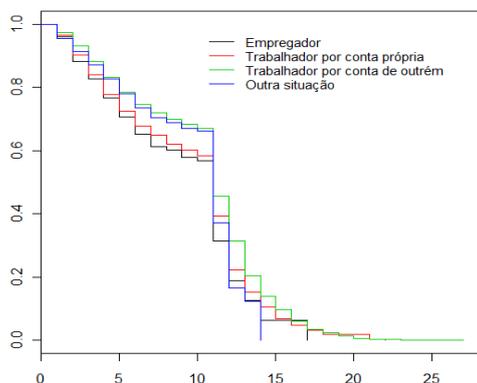


Figura 38b: Situação Profissional no ano 2000

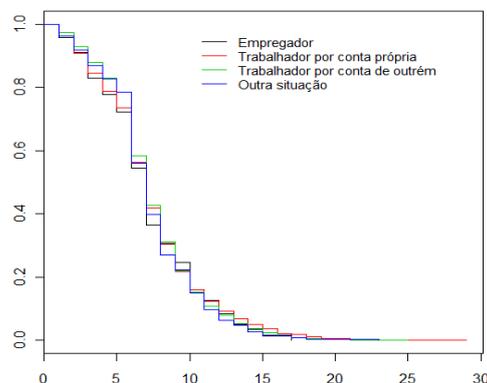
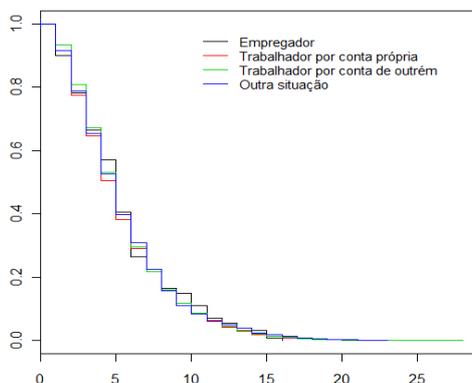


Figura 38c: Situação Profissional no ano 2005



O grupo de idade da mãe (figura 39) é uma covariável muito específica, já que à partida os comportamentos diferenciam-se de idade para idade no próprio ano. No entanto, o adiamento verificado nos últimos anos na fecundidade fez deste indicador um dos mais importantes na análise de fecundidade. Desta forma, os resultados obtidos, parecem corresponder à evolução da situação de fecundidade em Portugal. Ao observar o gráfico do ano de 1995 destaca-se o grupo de idades 13-25 como sendo o grupo que apresenta os níveis de probabilidade de ocorrer o evento mais elevados, nomeadamente, entre os 5 e 10 anos após o nascimento do primeiro filho. Porém, em 2000 e 2005, 5 e 10 anos depois, a situação é diferente. Esta alteração deve-se sobretudo ao aumento da idade ao nascimento do primeiro filho e à constante diminuição de fecundidade nas idades mais jovens. Ao comparar as figuras para os anos 1995, 2000 e 2005 verifica-se uma inversão de comportamentos: se em 1995 o grupo com probabilidades mais elevadas, ou

seja que esperava mais tempo até ao nascimento do segundo filho, era o grupo mais jovem, em 2005 a situação é totalmente oposta, este grupo passou a ser o que apresenta probabilidades menores e o grupo 33-54. Já os grupos de idade 26-28 e 29-32 anos apresentam situações não muito diferentes ao longo dos cinco anos, sendo que se observa, no entanto uma diminuição nas suas probabilidades.

**Figura 39:** Representação gráfica do grupo de idade da mãe nos anos de 1995, 2000 e 2005.

Figura 39a: Grupo de idade no ano 1995

Figura 39b: Grupo de idade no ano 2000

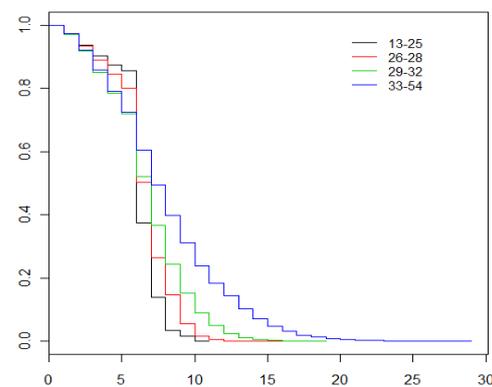
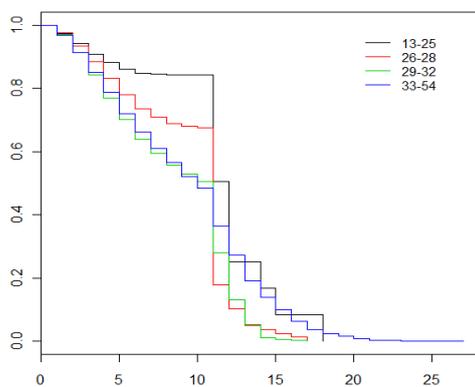
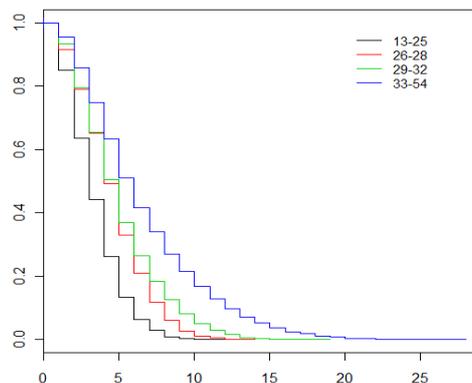


Figura 39c: Grupo de idade no ano 2005



Relativamente à covariável “filiação” (figura 40), não se verificam diferenças significativas entre os diferentes anos, sendo de distinguir, e como já foi referido anteriormente, que as probabilidades do tempo decorrido entre o nascimento do primeiro e do segundo filho, para as mães que têm os seus filhos fora do casamento é

superior às mães com filhos dentro do casamento. Apenas no ano de 2005 se observa nos primeiros anos que o tempo de espera é maior para as mães em casamento.

**Figura 40:** Representação gráfica da filiação nos anos de 1995, 2000 e 2005.

Figura 40a: Filiação no ano 1995

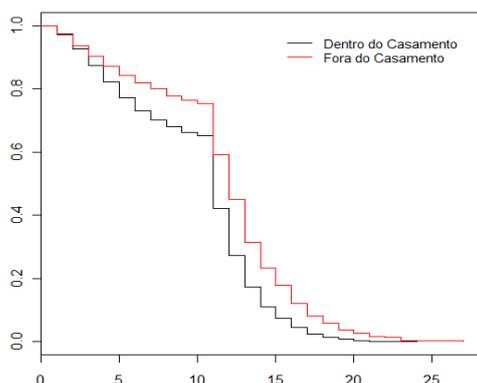


Figura 40b: Filiação no ano 2000

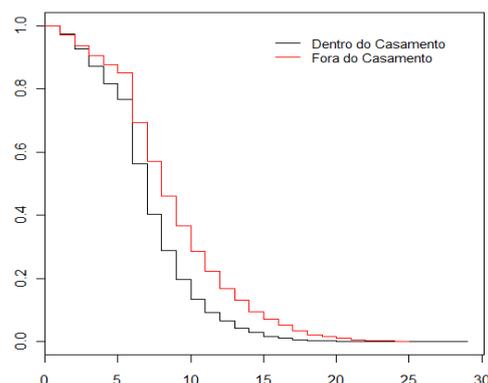
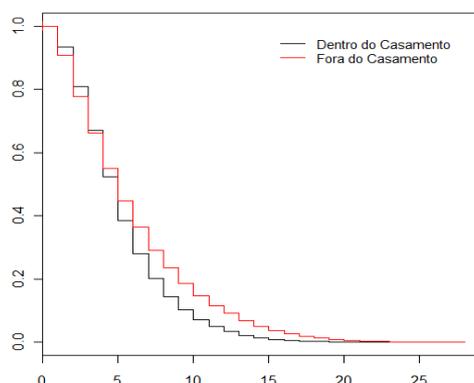


Figura 40c: Filiação no ano 2005



Após uma análise mais profunda das covariáveis utilizadas nos modelos, e com o objectivo de compreender o comportamento evolutivo entre o nascimento do primeiro e do segundo filho em Portugal para o período de 1995 a 2005, procede-se à apresentação dos modelos finais de acordo com cada um dos anos em análise. Desta forma na tabela 9, encontra-se a síntese dos modelos finais para cada ano em análise com os seus respectivos valores de *HR* e onde à semelhança do que aconteceu na análise do espaçamento entre o casamento e o nascimento do primeiro filho, se distinguem diferentes comportamentos.

Os modelos finais para cada ano em análise apresentam semelhanças nas variáveis explicativas na evolução do espaçamento entre o nascimento do primeiro e segundo filho. No período em análise, destaca-se a importância das covariáveis filiação, categoria profissional da mãe e idade que se apresentam em todos os modelos finais., Para além disso, de referir também a covariável “instrução da mãe” presente em todos os anos com exceção de 2001 e a covariável “condição perante o trabalho da mãe” presente somente no ano de 2004. No que diz respeito às interações entre as covariáveis é importante referir os seus comportamentos; observa-se a interação entre as covariáveis: filiação e instrução da mãe em todos os anos em análise à exceção de 1996; filiação e categoria profissional da mãe à exceção dos anos de 2001 e 2002; filiação e grupo de idade da mãe para os anos de 1995, 1996, 1999 e 2003; nível de instrução e grupo de idade da mãe nos anos de 1995, 1996, 2002 e 2005; e, finalmente, a interação entre a categoria profissional e a idade da mãe à exceção do ano de 2002.

Uma análise dos resultados obtidos anualmente é bastante importante. Todavia, e ainda que semelhantes nos seus resultados os modelos não são iguais. Porém, uma análise evolutiva de algumas covariáveis é interessante do ponto de vista demográfico. Assim, e tomando como exemplo a covariável “filiação” e observando a sua evolução entre 1995 e 2005, verificamos que em 1995 o tempo decorrido entre o nascimento do primeiro e do segundo filho era para as mães que tinham os seus filhos fora do casamento de -18%, isto é, se assumirmos probabilidade = tempo, este valor significava que as mães com filhos fora do casamento tinham de esperar mais tempo até ao nascimento do segundo filho. Já em 2005, esta situação inverteu-se e a probabilidade para o tempo de espera aumentou duas vezes e meia para as mães com filhos dentro do casamento, ou seja, as mães com filhos fora do casamento esperam menos tempo que as mulheres casadas para o nascimento do segundo filho.

**Tabela 9:** Valores de *HR* dos modelos finais de por cada ano em análise

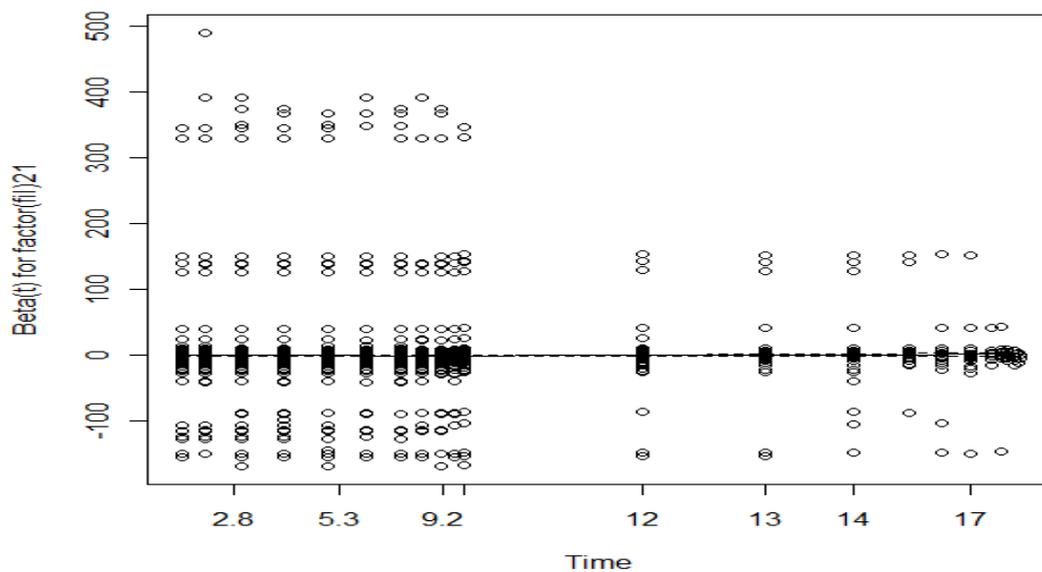
	1995	1996	1999	2000	2001	2002	2003	2004	2005
fil21	0,82	0 5	0,73	1,28	0,46	1,38	2,03	2,19	2,54
instrmae_categ1	0,63	0,65	0,9	0,87	-	0,86	1,04	1,07	2,19
instrmae_categ2	0,65	0,73	0,88	0,92	-	0,02	1,32	1,35	3,02
instrmae_categ3	1,09	1,12	1,01	1,04	-	1,57	1,88	1,98	4,86
cdtbmae3	-	-	-	-	-	-	-	0,57	-
profmae_categ1	0,73	0,78	0,64	0,65	0,58	0,9	0,85	0,87	2,87
profmae_categ2	0,62	0,72	0,56	0,61	0,52	0,91	0,88	0,91	1,78
idademae_categ21	1,1	1,14	0,49	0,57	0,67	1,17	1,23	1,67	1,86
idademae_categ22	1,21	1,37	1,17	1,23	1,31	0,62	1,59	1,72	1,62
fil21:instrmae_categ1	0,85	-	0,72	0,61	1,54	0,6	0,45	0,39	0,42
fil21:instrmae_categ2	0,68	-	0,68	0,47	2	0,49	0,37	0,34	0,33
fil21:instrmae_categ3	0,51	-	0,6	0,38	0,92	0,38	0,27	0,28	0,32
fil21:profmae_categ1	1,14	1,39	1,25	1,06	0,75		0,9	0,94	0,78
fil21:profmae_categ2	1,26	1,77	1,53	0,79	-		0,77	0,91	0,76
fil21:idademae_categ21	1,05	1,05	1,05	-	-	0,96	-	-	-
fil21:idademae_categ22	0,73	0,68	0,78	-	-	0,85	-	-	-
cdtbmae3:idademae_categ21	-	-	-	-	-	-	-	0,72	-
cdtbmae3:idademae_categ22	-	-	-	-	-	-	-	0,79	-
instrmae_categ1:cdtbmae3	-	-	-	-	-	-	-	-	-
instrmae_categ2:cdtbmae3	-	-	-	-	-	-	-	-	-
instrmae_categ3:cdtbmae3	-	-	-	-	-	-	-	-	-
instrmae_categ1:idademae_categ21	1,44	1,44	-	-	-	1,4	-	-	0,33
instrmae_categ2:idademae_categ21	0,92	0,89	-	-	-	1,23	-	-	0,32
instrmae_categ3:idademae_categ21	0,36	0,37	-	-	-	0,69	-	-	0,3
instrmae_categ1:idademae_categ22	1,54	1,54	-	-	-	2,07	-	-	0,55
instrmae_categ2:idademae_categ22	1,46	1,29	-	-	-	2,55	-	-	0,51
instrmae_categ3:idademae_categ22	0,81	0,79	-	-	-	2,24	-	-	0,24
profmae_categ1:idademae_categ21	1,01	0,98	2,41	2,23	2	-	1,31	1,07	0,99
profmae_categ2:idademae_categ21	1,47	1,36	3,59	3,07	2,5	-	1,28	0,95	0,89
profmae_categ1:idademae_categ22	1,29	1,19	1,42	1,31	1,16	-	0,92	0,83	0,89
profmae_categ2:idademae_categ22	1,68	1,31	1,53	1,28	1,05	-	0,77	0,68	0,75

Os resultados obtidos nestes modelos, que se apresentam como os modelos explicativos e parcimoniosos para a análise do espaçamento entre o nascimento do primeiro e do segundo filho, do ponto de vista do contexto demográfico da evolução da fecundidade apresentam-se significativos. No entanto, e tendo em conta a aplicação de uma abordagem estatística que caracteriza este trabalho, é importante verificar se estes modelos apresentados satisfazem os pressupostos estatísticos associados à análise de sobrevivência e nomeadamente aos modelos proporcionais de Cox.

Apresentamos assim a representação gráfica associada aos pressupostos do modelo de Cox para o ano de 1995, isto tendo em conta o número de covariáveis (7) em análise, bem como o número de anos em análise (9 anos).

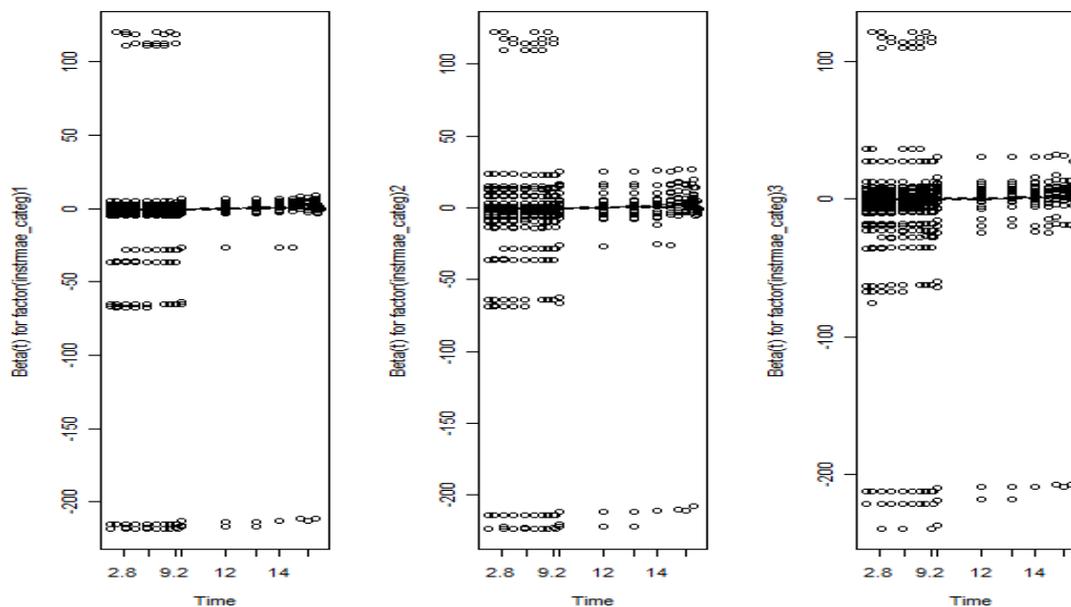
Relativamente aos resíduos de Schoenfeld padronizados, apresentamos os resultados para as covariáveis “filiação” e “nível de instrução da mãe”. Na figura 41 pode observar-se o comportamento para resíduos de Schoenfeld padronizados da covariável “filiação”, verificando que a distribuição não é de forma alguma aleatória, identificando-se assim uma tendência na distribuição dos resultados.

**Figura 41:** Resíduos de Schoenfeld padronizados para a covariável filiação no ano de 1995



Relativamente à representação dos resíduos de Schoenfeld padronizados para a covariável “instrução da mãe” (figura 42) verifica-se, uma vez mais, uma distribuição dos resultados com um padrão não aleatório dos dados em nenhuma das categorias desta covariável.

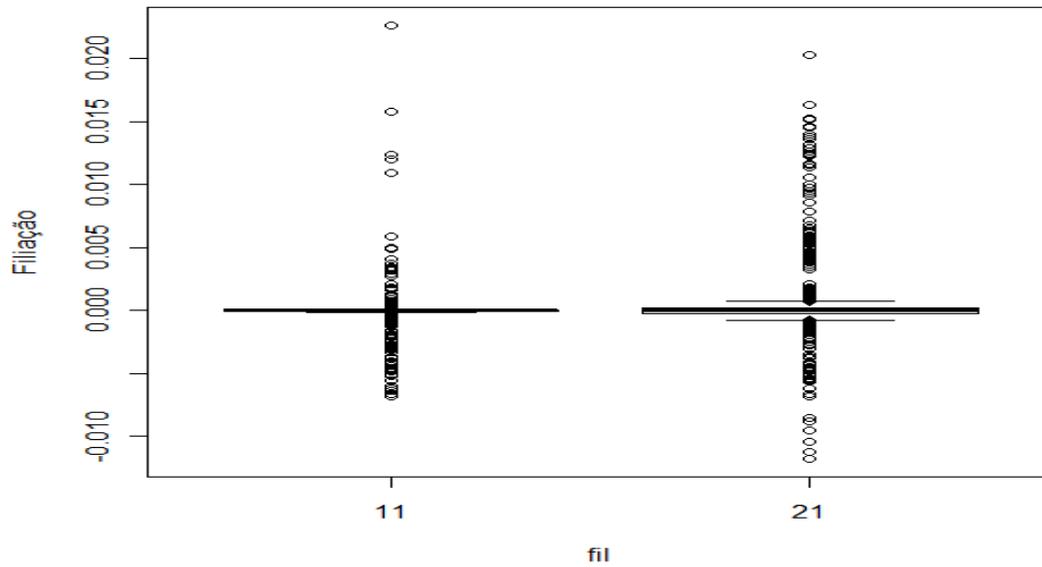
**Figura 42:** Resíduos de Schoenfeld padronizados para a covariável instrução da mãe no ano de 1995



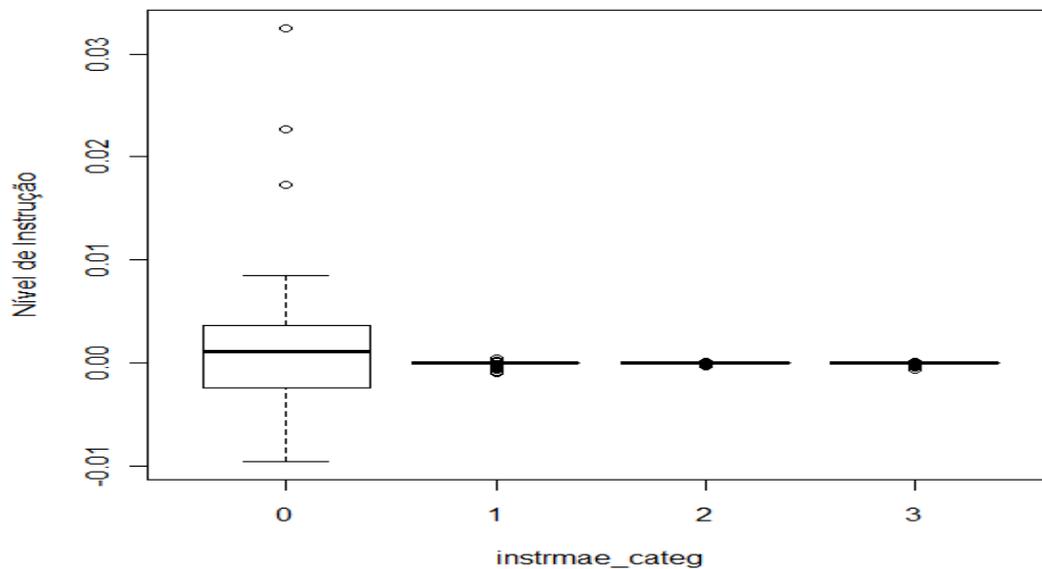
Como acontecia no espaçamento entre o casamento e o nascimento do primeiro filho, também aqui a não verificação dos riscos proporcionais nas covariáveis apresentadas através dos resíduos de schoenfeld padronizados, se verificou nas restantes covariáveis, acentuando uma vez mais a hipótese de diferentes abordagens a ter em conta, no futuro.

É, todavia, importante observar os resíduos *score*, para estas mesmas covariáveis permitindo a verificação da influência de cada observação no ajuste do modelo. Na representação gráfica dos resíduos *score* observa-se para a covariável “filiação” (figura 43) a presença de inúmeras observações influentes que poderão influenciar os resultados finais obtidos. Porém, para a covariável “instrução da mãe” (figura 44) observa-se que apenas a primeira categoria (sem instrução) parece apresentar observações influentes ou outliers.

**Figura 43:** Resíduos Score para a covariável filiação no ano de 1995



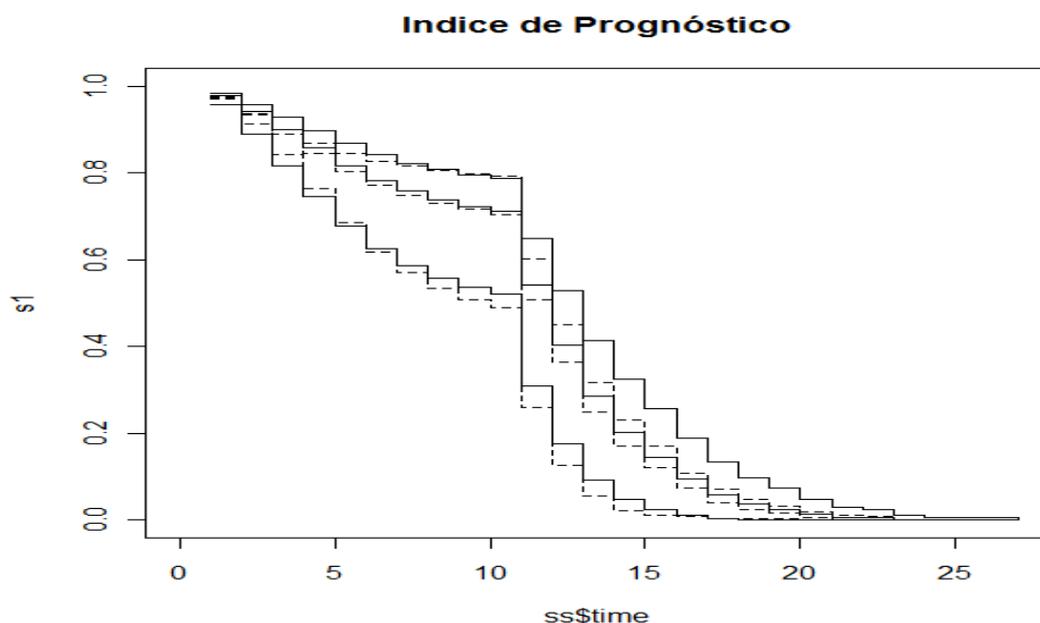
**Figura 44:** Resíduos Score para a covariável instrução da mãe no ano de 1995



Ainda que as covariáveis utilizadas, para explicar a evolução no espaçamento entre o nascimento do primeiro e do segundo filho, pareçam apresentar resultados não satisfatórios para se considerarem modelos de riscos proporcionais de Cox, é ainda relevante realizar um teste de índice de prognóstico de forma a verificar a

adequabilidade dos modelos finais. A representação gráfica do índice de prognóstico (figura 45) possibilita a observação nas linhas sólidas o modelo final ajustado e nas linhas tracejadas as estimativas Kaplan-Meier. Assim, e embora os modelos finais pareçam não assumir os riscos proporcionais, ao observar o índice de prognóstico, verificamos que parece de facto existir um bom ajustamento dos modelos.

**Figura 45:** Índice de Prognóstico para o modelo de 1995



De acordo com os resultados obtidos nos pressupostos associados ao modelo de riscos proporcionais de Cox, não foram obtidos resultados significativos para afirmar a proporcionalidade dos modelos. Contudo, o resultado obtido com o índice de prognóstico é revelador de que os modelos são de facto explicativos da evolução do espaçamento entre o nascimento do primeiro e segundo filho.

Na evolução do espaçamento entre o nascimento do primeiro e do segundo filho, concluímos assim que no período em análise, entre 1995 e 2005, as covariáveis que influenciam o espaçamento e a decisão de ter o segundo filho são: a filiação do nascimento, a instrução, a categoria profissional e a idade da mãe. Na covariável “filiação” observou-se que são as mães fora do casamento que apresentam maior espaçamento entre o nascimento do primeiro e segundo filho. Já na covariável “nível de instrução da mãe” podemos observar que são as mães com ensino superior as que

menos esperam pelo nascimento de um segundo filho. A covariável “categoria profissional” permitiu-nos compreender que as mães da categoria profissional 3 (onde se incluem mulheres com empregos em áreas como a agricultura ou outros empregos não qualificados), são as que apresentam probabilidades mais elevadas no tempo de espera entre o nascimento dos seus filhos. Relativamente ao grupo de idades, verifica-se que se no início do período, em 1995, eram as mães mais novas que apresentavam probabilidades mais elevadas, no final do período, em 2005, a situação inverte-se e são as mães com idades mais avançadas que apresentam probabilidades de tempo de espera mais elevadas.

### **3.2.1.3. Espaçamento entre o nascimento do segundo e o terceiro filho**

Nas últimas décadas, tal como referido anteriormente, a diminuição no número de filhos e o adiamento dos nascimentos tem sido questão fulcral na análise de fecundidade. É importante neste contexto compreender o comportamento das mães portuguesas e o que influencia o espaçamento entre o nascimento do seu segundo e terceiro filhos, sendo que muitas mães portuguesas adiam o nascimento dos seus filhos, nomeadamente do terceiro filho. Na aplicação de análise de sobrevivência os dados referentes aos anos de 2006 e 2007 foram truncados por se apresentarem anómalos. Para compreender a evolução no espaçamento entre o nascimento do segundo e do terceiro filho, as variáveis utilizadas foram:

- Instrução da Mãe,
- Condição perante o trabalho da Mãe,
- Categoria profissional da Mãe,
- Situação profissional da Mãe,
- Grupo de idade da Mãe,
- Filiação,
- Ano do nascimento (apenas numa análise global).

Como no caso anterior foram elaboradas recodificações de algumas variáveis, o ano do nascimento foi agrupado (1995/99, 2000/01, 2002/03 e 2004/05), assim como a idade da mãe (16-33, 34-36 e 37-42 anos).

A análise inicia-se, uma vez mais, pela validação individual das variáveis com um  $\alpha = 0,20$  (tabela 10). Ao observarmos os valores  $p$  verificamos que todas as covariáveis são consideradas altamente significativas para a construção de modelos; no entanto, a covariável relativa à “situação profissional das mães” não se apresenta significativa em alguns dos anos em análise (1995, 1999, 2002 e 2003), assim como a covariável “condição perante o trabalho” também se apresenta não significativa no ano de 2003.

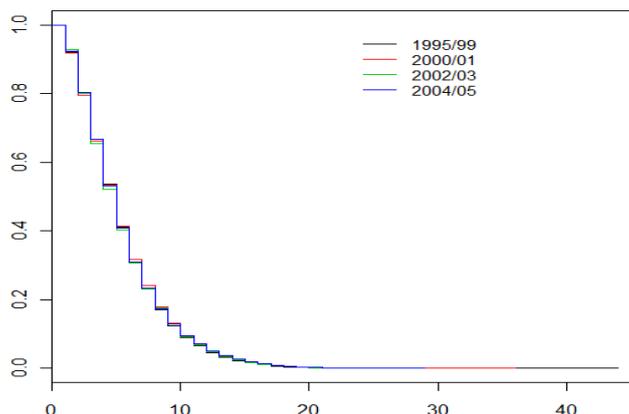
**Tabela 10:** Valores  $p$  para avaliar variáveis significativas a incluir nos modelos anuais

	Ano Nascimento Categorizado	Filiação	Nível de Instrução da Mãe	Condição perante o trabalho da Mãe	Profissão da Mãe	Situação Profissional da Mãe	Grupo de Idade da Mãe
Global	0,070	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
1995	-	<0.001	<0.001	0,020	<0.001	0,210	<0.001
1996	-	<0.001	<0.001	0,070	<0.001	0,050	<0.001
1999	-	<0.001	<0.001	0,020	<0.001	0,720	<0.001
2000	-	<0.001	<0.001	<0.001	<0.001	0,020	<0.001
2001	-	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
2002	-	<0.001	<0.001	0,030	<0.001	0,490	<0.001
2003	-	<0.001	<0.001	0,220	<0.001	0,390	<0.001
2004	-	<0.001	<0.001	<0.001	<0.001	0,010	<0.001
2005	-	<0.001	<0.001	0,100	<0.001	0,040	<0.001

Precedendo a apresentação dos modelos finais, é importante realizar uma breve análise gráfica da evolução das covariáveis apresentadas, de forma a compreender o comportamento de mães com o segundo filho e o espaçamento decorrido até ao nascimento do terceiro. Apresentam-se assim, primeiramente, os resultados para o período em análise de uma forma global e procederemos posteriormente a uma análise quinquenal das covariáveis (nos anos de 1995, 2000 e 2005).

No que diz respeito ao “ano do nascimento” (figura 46) e no contexto do espaçamento entre o segundo e terceiro filho, não se observam diferenças nos diferentes anos em análise, evidenciando-se que aproximadamente 60% das mães tem o seu terceiro filho até cinco anos após o nascimento do segundo filho.

**Figura 46** : Ano de nascimento do terceiro filho de mães portuguesas no período entre 1995 e 2005



Ao observar o “nível de instrução da mãe” (figura 47) no período em análise, verificamos que são as mães com um nível de ensino superior as que esperam menos tempo no que diz respeito ao espaçamento decorrido entre o nascimento do segundo e do terceiro filhos, enquanto em situação oposta estão as mães com ensino básico, revelando este comportamento uma situação diferente da observada anteriormente entre o espaçamento do casamento e o nascimento do primeiro filho, bem como entre o nascimento do primeiro e o segundo filho. No caso da “condição perante o trabalho da mãe” (figura 48) observa-se que as mães desempregadas têm uma probabilidade inferior às mães empregadas de terem o seu terceiro filho e são também as mães desempregadas que esperam mais tempo até ao nascimento deste filho.

**Figuras 47 e 48:** Representação gráfica da instrução da mãe e a sua condição perante o trabalho no período em análise.

Figura 47: Instrução da mãe

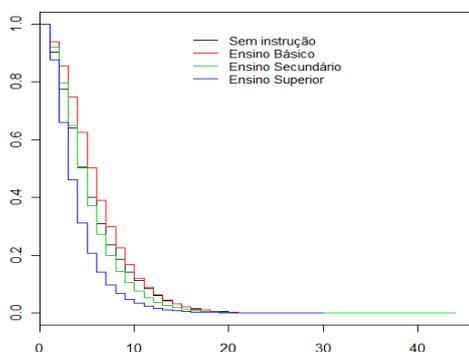
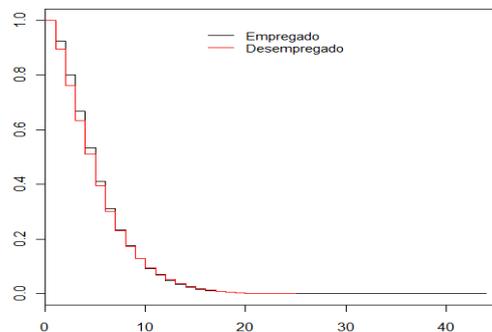


Figura 48 : Condição perante o trabalho



Ao observarmos a “categoria profissional” (figura 49) verificamos que as mães da categoria profissional 3 são as que esperam mais tempo entre o nascimento do segundo e terceiro filhos, estando uma vez mais em situação oposta mulheres com a categoria profissional 1, categoria esta à qual se encontram associadas profissões exigentes de um nível de educação mais elevado. Já no que diz respeito à “situação profissional” (figura 50) não se observam diferenças entre as diferentes categorias destacando-se apenas que existe uma probabilidade de aproximadamente 50% para que as mulheres tenham o seu terceiro filho até 5 anos após o nascimento do segundo filho.

**Figuras 49 e 50:** Representação gráfica da categoria profissional e situação profissional da mãe no período em análise.

Figura 49: Categoria Profissional

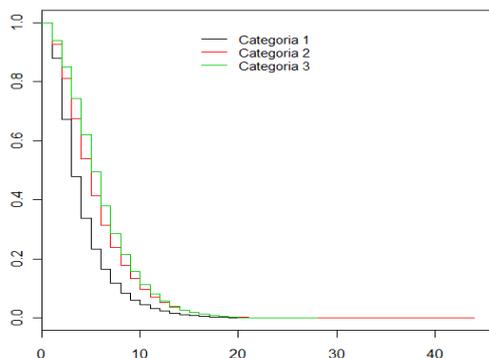
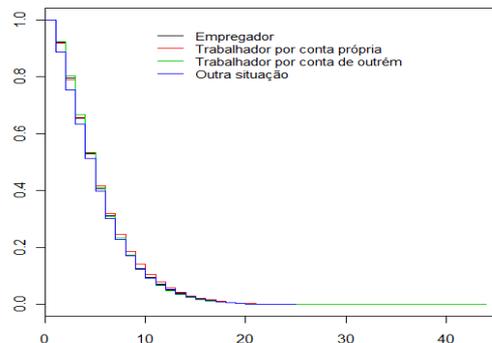


Figura 50: Situação Profissional

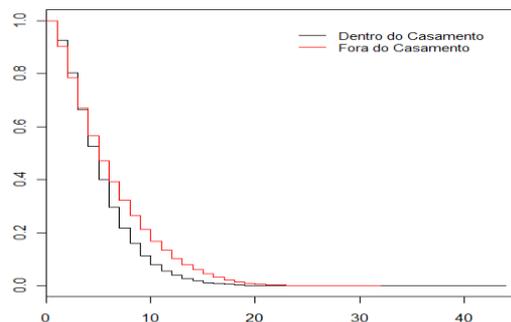
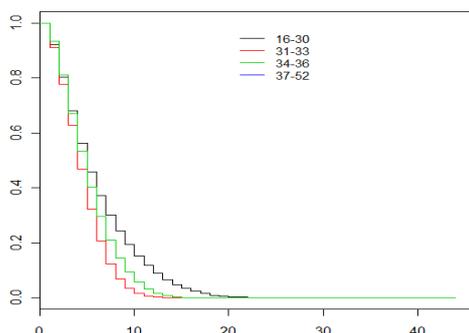


É importante uma vez reafirmar que a covariável grupo de “idade da mãe” (figura 51) apresenta características diferentes das restantes variáveis. Porém é de referir que existe uma probabilidade de cerca de 80% das mães, independentemente do grupo etário, terem o seu terceiro filho dois anos após o nascimento do segundo filho. Por outro lado, e no que diz respeito à “filiação” (figura 52), verifica-se que até 5 anos depois do nascimento do segundo filho a probabilidade do nascimento de um terceiro filho é semelhante dentro e fora do casamento. Todavia, após este período as mães com filhos fora do casamento esperam mais tempo até ao nascimento de um terceiro filho.

**Figuras 51 e 52:** Representação gráfica do grupo de idade da mãe e da filiação no período em análise.

Figura 51: Grupo de Idade da Mãe

Figura 52: Filiação



Após uma apresentação do comportamento das covariáveis explicativas da evolução do espaçamento entre o nascimento do segundo e terceiro filhos, em Portugal, no período 1995 a 2005, é necessário realizar uma abordagem de modelo de Cox estratificado. Como realizado anteriormente na análise do comportamento entre o casamento e o nascimento do primeiro filho, e para o espaçamento entre o nascimento do primeiro e segundo filho, apresentamos em seguida a evolução quinquenal para cada uma das variáveis explicativas dos modelos - nível de instrução, condição perante o trabalho, categoria profissional, situação profissional, grupo de idade da mãe e filiação do nascimento.

Ao observar a evolução do “nível de instrução” (figura 53) para os anos de 1995, 2000 e 2005 verificamos que não existem diferenças significativas nestes três anos; salienta-se como anteriormente que são as mães com ensino superior que menos tempo esperam até ao nascimento do seu terceiro filho em oposição às mulheres com o ensino básico.

**Figura 53:** Representação gráfica do nível de instrução da mãe nos anos de 1995, 2000 e 2005.

Figura 53a: Nível de instrução no ano 1995

Figura 53b: Nível de instrução no ano 2000

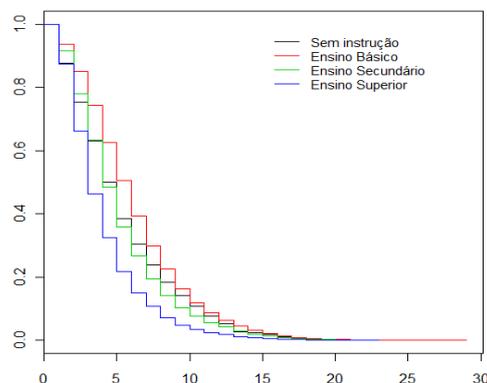
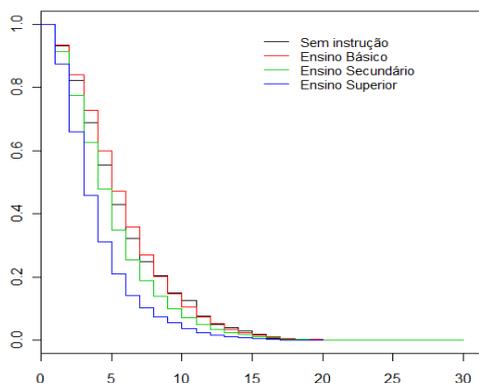
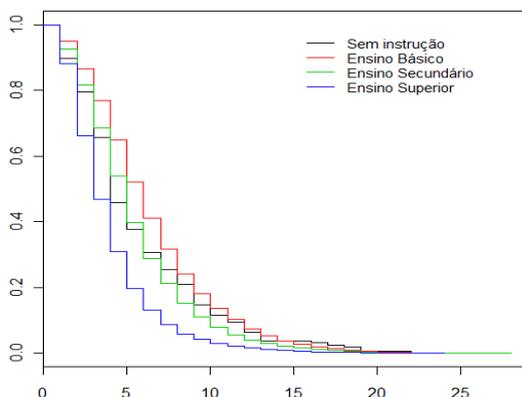


Figura 53c: Nível de instrução no ano 2005



Na observação da “condição perante o trabalho da mãe” (figura 54) não se observam diferenças significativas entre os diferentes anos; no entanto, no ano de 2000 verifica-se um ligeiro aumento no espaçamento entre o nascimento do segundo e terceiro filho, nomeadamente, para as mulheres empregadas. Já no ano de 2005, não existem de facto diferenças significativas, observando-se apenas um “tempo de espera” mais elevado para as mulheres empregadas, quando passou um, dois ou três anos após o nascimento do segundo filho.

**Figura 54:** Representação gráfica da condição perante o trabalho da mãe nos anos de 1995, 2000 e 2005.

Figura 54a: Condição perante o trabalho no ano 1995

Figura 54b: Condição perante o trabalho no ano 2000

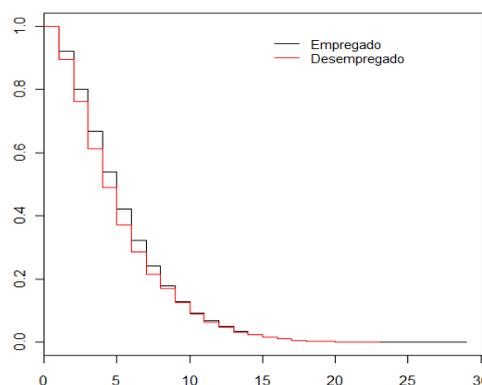
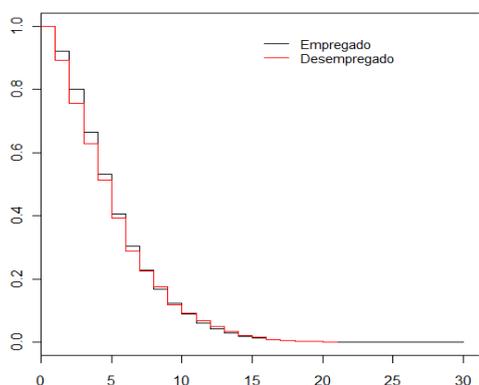
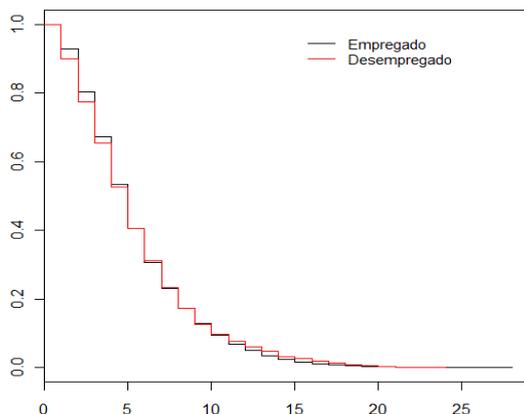


Figura 54c: Condição perante o trabalho no ano 2005



Relativamente à evolução da “categoria profissional” (figura 55), observa-se um distanciamento entre as categorias ao longo dos cinco anos em análise. Embora não se verifiquem diferenças substanciais entre o comportamento das mães pertencentes às diferentes categorias, verifica-se que, como referido anteriormente na análise do período, se observa um tempo de espera maior para as mães da categoria 3, enquanto as mães da categoria 1 são as que menos tempo esperam entre o nascimento do segundo e terceiro filho.

**Figura 55:** Representação gráfica da categoria profissional da mãe nos anos de 1995, 2000 e 2005.

Figura 55a: Categoria Profissional no ano 1995

Figura 55b: Categoria Profissional no ano 2000

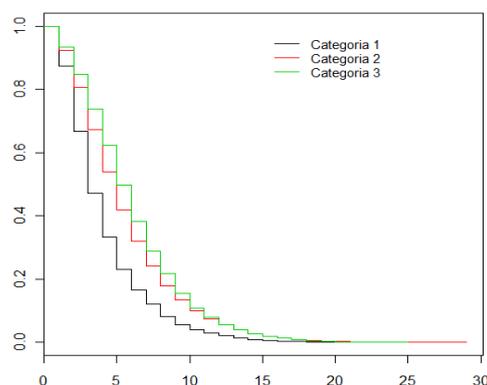
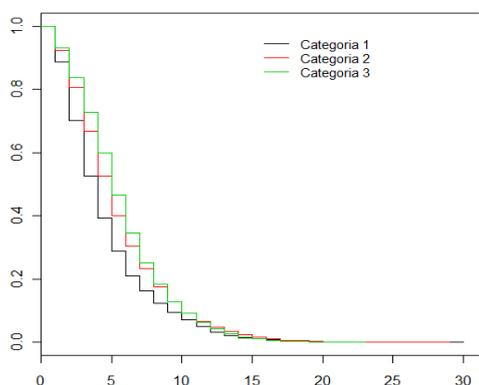
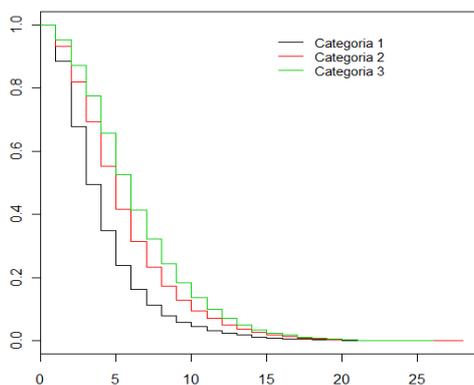


Figura 55c: Categoria Profissional no ano 2005



A “situação profissional” (figura 56) não apresenta diferenças entre os três anos em análise, apenas de referir que 50% das mães independentemente da sua situação profissional quando decide ter o seu terceiro filho, têm-no até cinco anos após o nascimento do segundo filho.

**Figura 56:** Representação gráfica da situação profissional da mãe nos anos de 1995, 2000 e 2005.

Figura 56a: Situação Profissional no ano 1995

Figura 56b: Situação Profissional no ano 2000

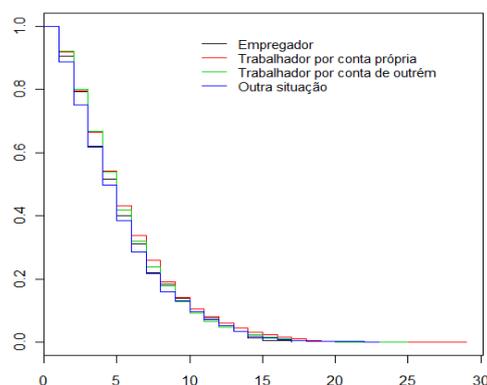
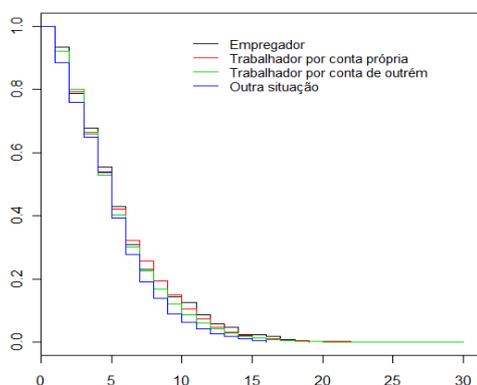
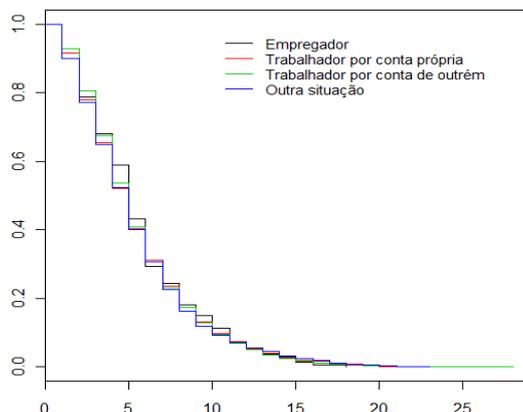


Figura 56c : Situação Profissional no ano 2005



O grupo de idade da mãe (figura 57) é uma das covariáveis mais importantes na análise dos comportamentos de fecundidade, e é por essa razão muito relevante analisar a evolução dos diferentes grupos de idade. Assim, refere-se o grupo de idade 34-36 anos, que se caracteriza por probabilidades de ocorrência do evento mais elevadas nos primeiros anos após o nascimento do segundo filho, diminuindo esta probabilidade cinco anos após o nascimento do segundo filho. Já os outros dois grupos de idade (14-33 e 37-52) apresentam comportamentos muito similares entre si, sendo que, pelas próprias características do grupo de idades, é o grupo 14-33 aquele em que as mães têm probabilidades mais elevadas de esperar mais tempo até ao nascimento do terceiro filho.

**Figura 57:** Representação gráfica do grupo de idade da mãe nos anos de 1995, 2000 e 2005.

Figura 57a: Grupo de idade no ano 1995

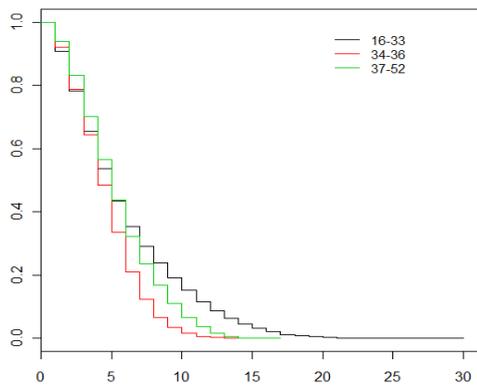


Figura 57b: Grupo de idade no ano 2000

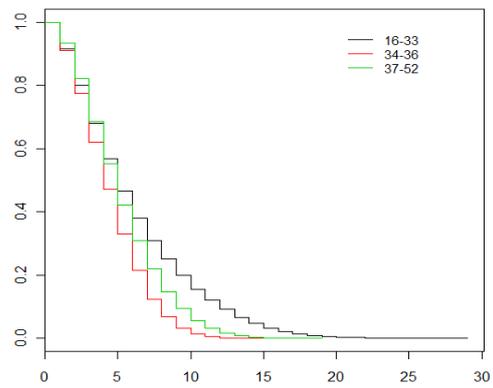
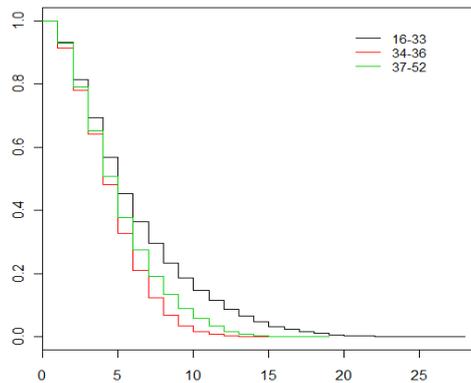


Figura 57c: Grupo de idade no ano 2005



No que diz respeito à covariável “filiação” (figura 58), verifica-se que não existem diferenças significativas nos anos em análise, sendo que são as mães fora do casamento que mais tempo têm de esperar até ao nascimento do terceiro filho quando o decidem ter.

**Figura 58:** Representação gráfica da filiação nos anos de 1995, 2000 e 2005.

Figura 58a: Filiação no ano 1995

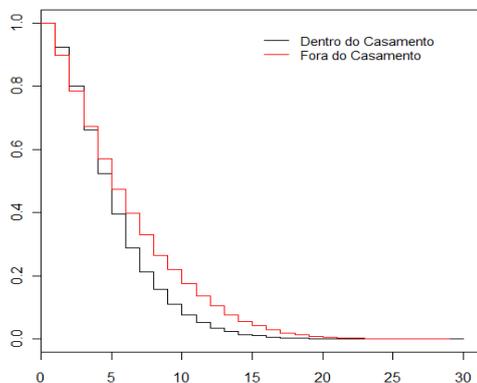


Figura 58b: Filiação no ano 2000

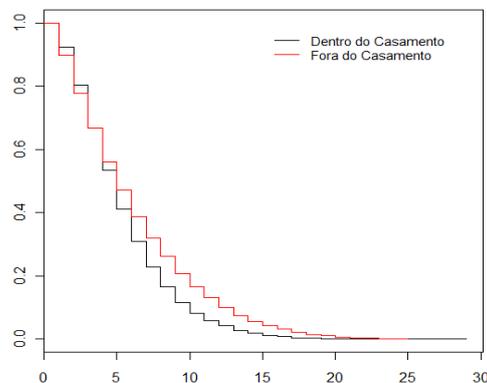
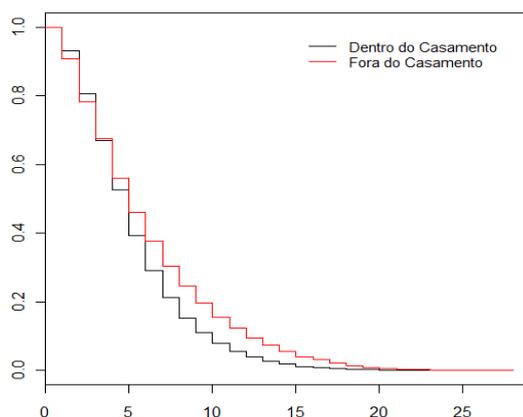


Figura 58c: Filiação no ano 2005



Os resultados apresentados pelos *HR* reflectem os comportamentos observados anteriormente aquando da análise gráfica das covariáveis (tabela 11). Tendo por base o ano de 2005, observemos duas das covariáveis, a filiação e a instrução da mãe. Ao observar a covariável “filiação” verificamos que para as mães fora do casamento a probabilidade de terem o seu terceiro filhos é 74% mais que para as mães casadas. Já no caso da “instrução da mãe” observa-se que existe uma probabilidade de mais 19% para as mães com ensino básico (categoria 1) relativamente às mães sem instrução (categoria 2), o mesmo acontece às mães com ensino secundário (categoria 2) que apresentam uma probabilidade de 56% maior que as mães na categoria de referência. As mães com o

ensino superior (categoria 3) apresentam também valores elevados sendo que a probabilidade de estas terem o terceiro filho relativamente às mães da categoria de referência é de aproximadamente duas vezes e meia a mais.

**Tabela 11:** Valores de *HR* dos modelos finais de por cada ano em análise

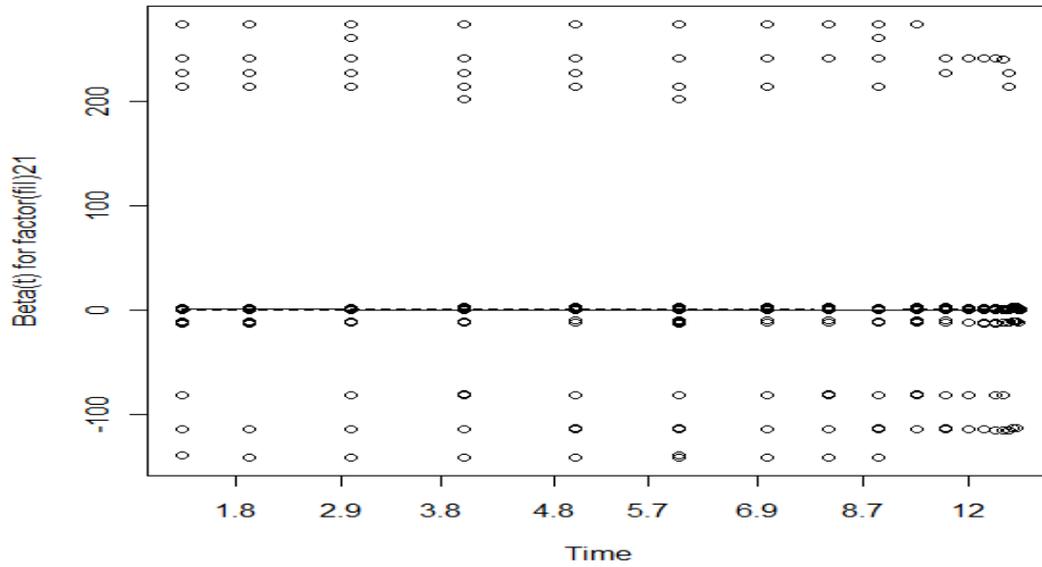
	19 5	1996	1999	2000	2001	2002	2003	2004	2005
fil21	1,25	1,21	1,06	1,48	1,41	1,42	2,31	2,01	1,74
instrmae_categ1	1,01	0,85	0,97	0,99	1,11	0,47	0,67	1,16	1,19
instrmae_categ2	1,09	1,14	1,24	1,36	,48	0,63	0,95	1,51	1,56
instrmae_categ3	1,51	1,65	1,79	1,88	2,16	0,93	1,51	2,22	2,38
cdtbmae	1,19	0,72	1,05	1,30	1,28	0,27	-	1,38	1,22
profmae_categ1	-	-	1,03	0,97	0,98	0,47	0,50	0,90	0,90
profmae_categ2	-	-	1,15	1,05	1,06	0,61	0,65	0,92	0,88
idademae_categ21	1,29	1,75	2,14	2,22	2,39	1,38	2,21	2,11	1,97
idademae_categ22	1,08	1,31	1,79	1,74	1,77	0,82	1,80	1,80	1,75
fil21:instrmae_categ1	0,66	0,75	0,74	0,65	0,61	0,64	0,40	0,44	0,57
fil21:instrmae_categ2	0,57	0,62	0,72	0,51	0,57	0,55	0,35	0,38	0,46
fil21:instrmae_categ3	0,49	0,46	0,62	0,47	0,47	0,47	0,27	0,33	0,44
fil21:cdtbmae3	1,15	1,21	1,22	-	1,15	1,17	-	-	-
fil21:profmae_categ1	-	-	1,04	-	1,04	-	-	-	-
fil21:profmae_categ2	-	-	1,26	-	1,18	-	-	-	-
fil21:idademae_categ21	-	0,86	0,84	0,88	0,84	0,92	0,92	0,93	0,88
fil21:idademae_categ22	-	0,81	0,85	0,86	0,81	0,90	0,83	0,93	0,83
cdtbmae3:idademae_categ21	0,86		-	-	0,78	-	-	0,78	0,77
cdtbmae3:idademae_categ22	0,75		-	-	0,82	-	-	0,82	0,67

Do ponto de vista demográfico, os resultados obtidos nos modelos previamente apresentados reflectem a evolução do espaçamento entre o nascimento do segundo e terceiro filho, para o período 1995 a 2005, em Portugal. Contudo, e reafirmando uma vez mais, é importante verificar se os modelos apresentados satisfazem os pressupostos de riscos proporcionais do modelo de Cox.

Como nas duas situações anteriores, apresentamos a representação gráfica associada aos pressupostos do modelo de Cox para o ano de 1995.

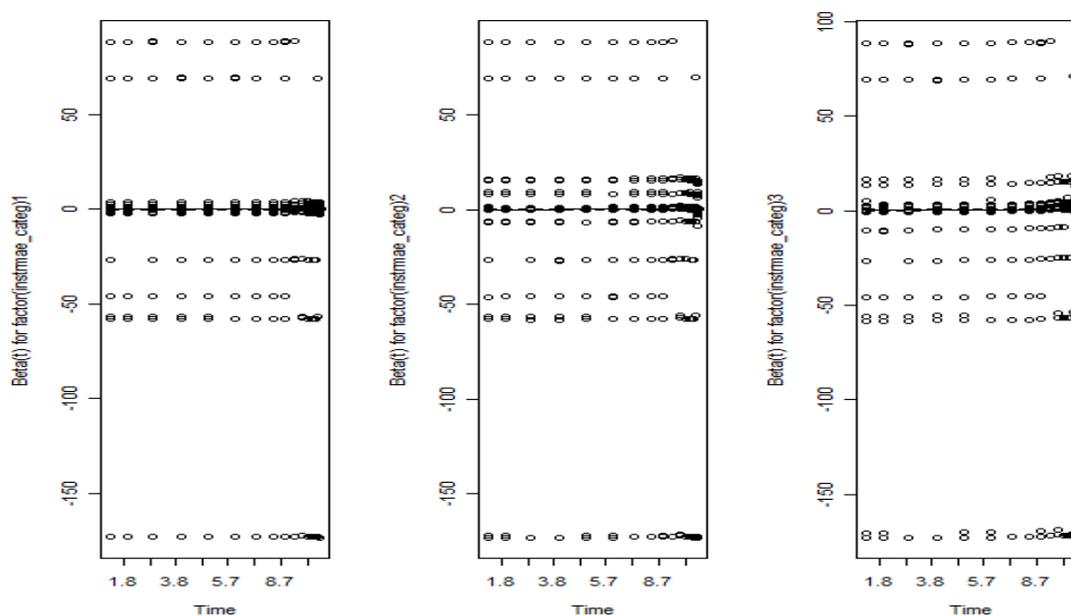
Relativamente aos resíduos de Schoenfeld padronizados, apresentamos os resultados para as covariáveis “filiação” e “nível de instrução” da mãe. Na figura 59 observa-se que o comportamento para resíduos de Schoenfeld padronizados da covariável “filiação” não apresenta uma distribuição aleatória.

**Figura 59:** Resíduos de Schoenfeld padronizados para a covariável filiação no ano de 1995



Relativamente à representação dos resíduos de Schoenfeld padronizados para a covariável “instrução da mãe” (figura 60) verifica-se uma vez mais a existência de uma distribuição dos resultados com um padrão não aleatório dos dados em nenhuma nas categorias desta covariável.

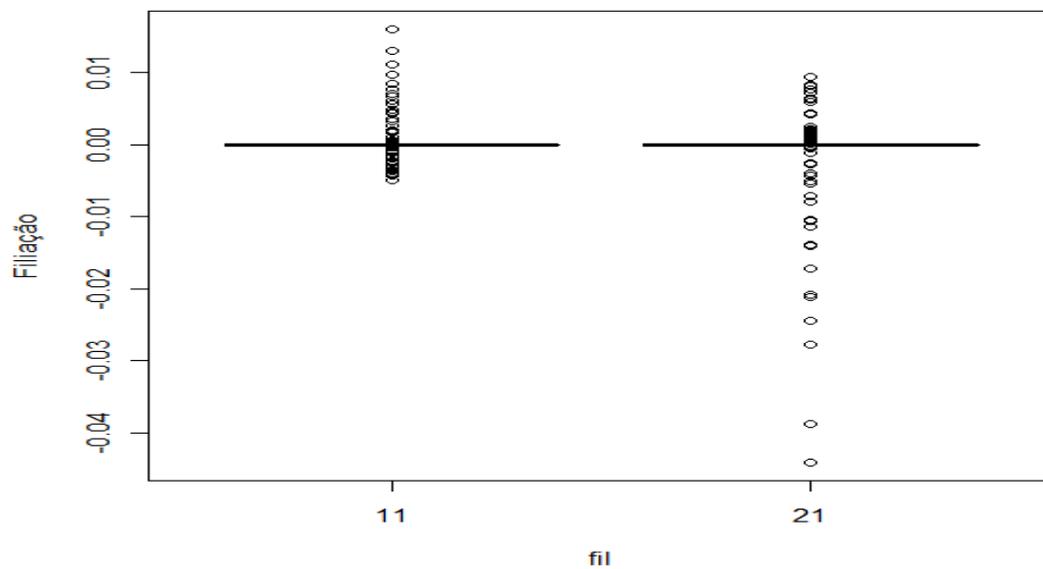
**Figura 60:** Resíduos de Schoenfeld padronizados para a covariável instrução da mãe no ano de 1995



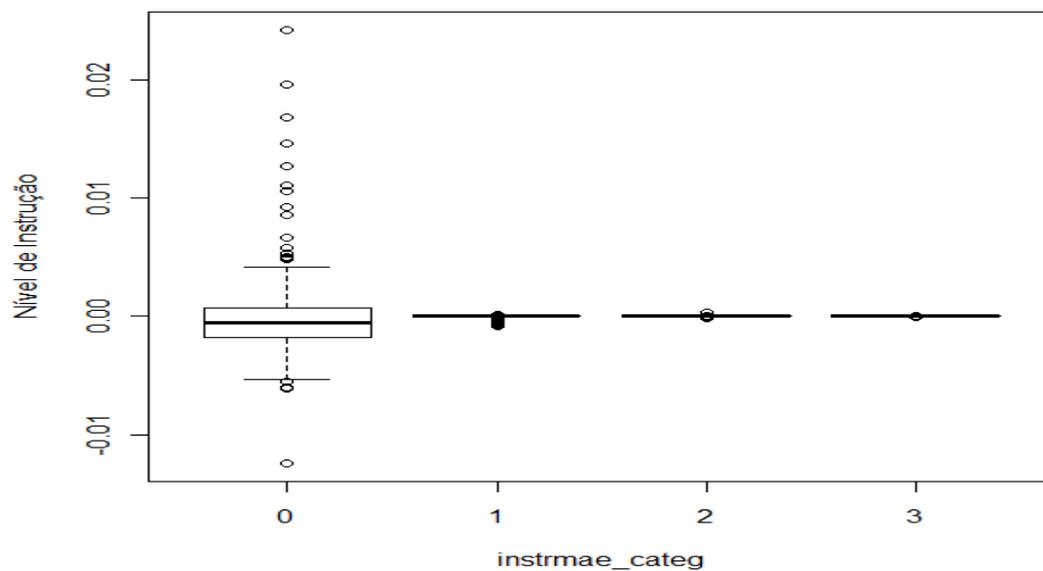
Como acontecia no espaçamento entre o casamento e o nascimento do primeiro filho, e também no espaçamento entre o nascimento do primeiro e segundo filho, na análise do espaçamento ente o nascimento do segundo e terceiro filho parecem não se verificar os pressupostos associados aos riscos proporcionais nas covariáveis apresentadas através dos resíduos de schoenfeld padronizados.

É contudo importante, e à semelhança das situações anteriores, observar a representação gráfica dos resíduos *score*, para estas mesmas covariáveis permitindo a verificação da influência de cada observação no ajuste do modelo. Assim, na representação gráfica dos resíduos *score* observa-se para a covariável “filiação” (figura 61) a presença de diferentes observações influentes que poderão influenciar os resultados finais obtidos, observando-se o mesmo comportamento na covariável “instrução da mãe” (figura 62).

**Figura 61:** Resíduos Score para a covariável filiação no ano de 1995

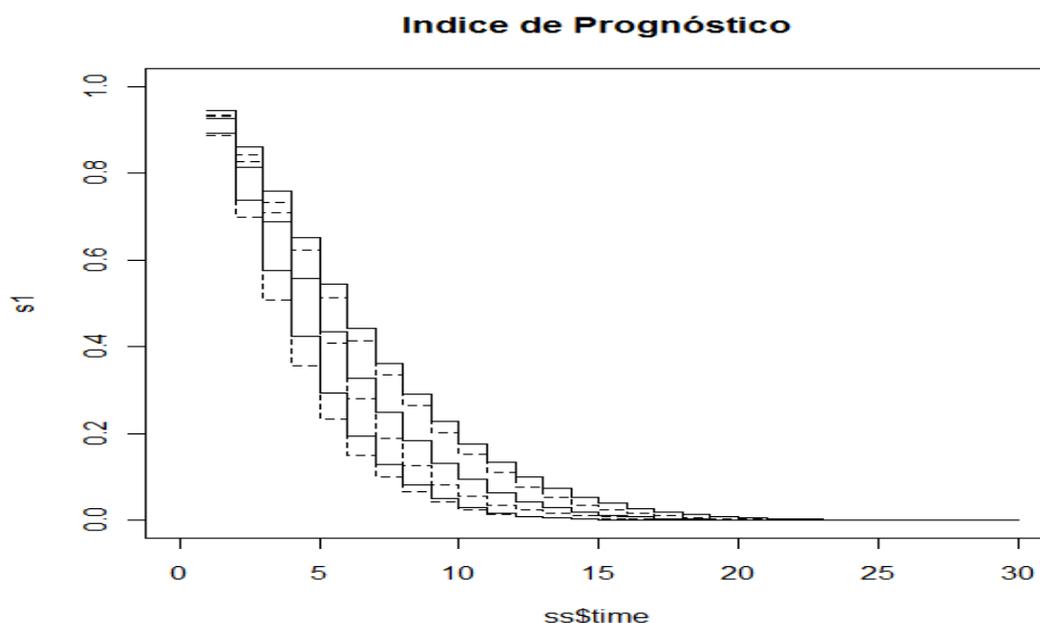


**Figura 62:** Resíduos Score para a covariável instrução da mãe no ano de 1995



Ainda que as covariáveis utilizadas para explicar a evolução no espaçamento entre o nascimento do segundo e do terceiro filho, pareçam ter apresentado resultados não satisfatórios para se considerar que os modelos de Cox apresentados seguem os riscos proporcionais, é ainda importante realizar um teste de índice de prognóstico de forma a verificar a adequabilidade dos modelos finais. O índice de prognóstico (figura 63) possibilita a observação nas linhas sólidas do modelo final ajustado e nas linhas tracejadas das estimativas Kaplan-Meier. Assim, e embora os modelos finais pareçam não assumir os riscos proporcionais, ao observar o índice de prognóstico verificamos que parece de facto existir um bom ajustamento dos modelos, já que as duas linhas seguem o mesmo comportamento e se sobrepõem por diversas vezes.

**Figura 63:** Índice de Prognóstico para o modelo de 1995



Tendo por base os resultados obtidos nos pressupostos associados ao modelo de riscos proporcionais de Cox, não foram obtidos resultados significativos para afirmar a proporcionalidade dos modelos. No entanto, o resultado obtido com o índice de prognóstico é revelador de que os modelos são de facto explicativos da evolução do espaçamento entre o nascimento do segundo e terceiro filhos.

Quanto a esta evolução, concluímos que no período em análise, entre 1995 e 2005, as covariáveis que influenciam o espaçamento e a decisão de ter o terceiro filho são a filiação do nascimento, a instrução, a condição perante o trabalho, a categoria profissional e a idade da mãe. Na covariável “filiação” observou-se que cinco anos após o nascimento do segundo filho, e no que diz respeito ao espaçamento até ao nascimento do terceiro filho, são as mães fora do casamento que apresentam um tempo de espera maior. Já na covariável “nível de instrução da mãe” podemos observar que são as mães com ensino superior as que menos esperam pelo nascimento de um terceiro filho. A covariável “condição perante o trabalho” revelou que após o nascimento do segundo filho não existem diferenças no que diz respeito à decisão e tempo decorrido até ao nascimento do terceiro filho. Por outro lado, a covariável “categoria profissional” permitiu-nos compreender que as mães da categoria profissional 3 (onde se incluem mulheres com empregos em áreas como a agricultura ou empregos não qualificados), apresentam probabilidade mais elevada no tempo de espera entre o nascimento dos seus filhos, estando em sentido oposto as mães da categoria profissional 1. Relativamente ao grupo de idades, verifica-se que são as mães mais novas seguidas das mais velhas que esperam mais tempo até ao nascimento do terceiro filho.

### **3.2.2. Aplicação do modelo aditivo de Aalen**

Foi também aplicado aos dados disponíveis para análise e nos três momentos de interesse, o espaçamento entre o casamento e o nascimento do primeiro filho, o espaçamento entre o nascimento do primeiro e do segundo filho e por fim no espaçamento decorrido entre o nascimento de segundo e o terceiro filho o modelo aditivo de Aalen. No entanto o resultado obtido foi nulo, devido à dimensão dos dados em análise, obtendo-se em qualquer um dos casos duas mensagens, “*cannot allocate vector of length*” e ainda “*too many elements specified*”. Desta forma não foi possível a aplicação do método.

#### 4. Modelos Lineares Generalizados

Os modelos lineares generalizados (MLG) surgem com o objectivo de analisar a influência de uma ou mais co-variáveis explicativas, sobre uma covariável de interesse. Os MLG são uma extensão do modelo linear clássico que propõe duas direcções diferentes, (1) por um lado a distribuição considerada não tem de ser normal, sendo contudo qualquer distribuição da família exponencial; (2) por outro lado, a função que relaciona o valor esperado e o vector de co-variáveis pode ser qualquer função diferenciável.

Os MLG pelo grande número de modelos que englobam, bem como pela facilidade de rápido desenvolvimento computacional, têm vindo a desempenhar um papel cada vez mais importante na análise estatística, ainda que com algumas limitações impostas, principalmente, na linearidade da estrutura, bem como, a restrição de se enquadrar na família exponencial e por exigirem também a independência das respostas.

##### 4.1. Metodologia

Os modelos lineares generalizados apresentam-se com uma estrutura de regressão linear e apresentam-se como sendo parte de uma família de distribuição com propriedades muito específicas: a distribuição exponencial. Uma covariável pertence à família exponencial quando a sua função de densidade de probabilidade se escreve:

$$f(y|\theta, \phi) = \exp\left\{\frac{y\theta - b\theta}{a(\phi)} + c(y, \phi)\right\} \quad [4.1]$$

onde  $\theta$  e  $\phi$  são parâmetros escalares,  $a(\cdot)$ ,  $b(\cdot)$  e  $c(\cdot, \cdot)$  são funções reais conhecidas.

É importante reforçar a ideia de que os MLG são uma extensão do modelo linear clássico:

$$Y = Z\beta + \varepsilon \quad [4.2]$$

onde  $Z$  é uma matriz de dimensão, associada a um vector  $\beta = (\beta_1, \dots, \beta_p)^T$  de parâmetros, e  $\varepsilon$  vector de erros aleatórios com uma distribuição que se assume como  $N_n(0, \sigma^2 I)$ . Implicando estas hipóteses que o valor esperado da covariável resposta é uma função linear das co-variáveis (Turkman e Silva 2000).

De acordo com Turkman e Silva, os MLG assumem uma distribuição que não tem necessariamente de ser normal, assumindo qualquer distribuição da família exponencial, e embora se mantenha a estrutura da linearidade, a função que relaciona o valor esperado e o vector de co-variáveis pode ser qualquer função diferenciável. Deste modo os MLG são caracterizados por (1) uma componente aleatória, onde dado o vector de co-variáveis  $x_i$ , as variáveis  $Y_i$  são independentes com distribuição enquadrada na família exponencial, com  $E(Y_i|x_i) = \mu_i = b'(\theta_i)$  para  $i = 1, \dots, n$  e com um parametro de dispersão  $\phi$  não dependente de  $i$ ; (2) uma componente estrutural ou sistemática, onde o valor esperado  $\mu_i$  está relacionado com o preditor linear  $\eta_i = z_i^T \beta$  através da relação:

$$\mu_i = h(\eta_i) = h(z_i^T \beta), \quad \eta_i = g(\mu_i) \quad [4.3]$$

onde,  $h$  é uma função monótona e diferenciável,  $g = h^{-1}$  é a função de ligação,  $\beta$  é um vector de parâmetros de dimensão  $p$  e  $z_i$  é um vector de especificação de dimensão  $p$  (função do vector de co-variáveis  $x_i$ ). De uma forma geral  $z_i = (1, x_{i1}, \dots, x_{ik})^T$  com  $k = p - 1$ . No entanto existem co-variáveis qualitativas que devem ser convenientemente codificada à custa de variáveis binárias consideradas variáveis *dummy*.

É ainda importante referir que os MLG enquadram em si alguns dos modelos mais importantes na análise estatística de dados, como podemos verificar na tabela 12.

**Tabela 12:** Síntese de alguns modelos lineares generalizados

Componente aleatória	Componente estrutural		Modelo
	Função de ligação	Covariáveis	
Normal	Identidade	Contínuas	Regressão linear
Normal	Identidade	Categorizadas	Análise de variância
Normal	Identidade	Mistas	Análise de covariância
Binomial	Logit	Mistas	Regressão logística
Poisson	Logaritmica	Mistas	Log-linear

Fonte: Turkman e Silva (2000), (Tabela 1.3)

A modelação através dos MLG pressupõem diferentes etapas, onde se passa pela (1) formulação dos modelos, (2) ajustamento dos mesmos e (3) selecção e validação dos modelos para análise. A última etapa tem como objectivo identificar sub-modelos com um número moderado de parâmetros adequados aos dados, detectando discrepâncias entre os dados e os valores preditos, averiguando a existência de outliers ou/e observações influentes.

Os métodos inferenciais utilizados nos MLG pressupõem que o modelo esteja correctamente especificado; todavia, existem excepções em que se verificam situações de sobre dispersão, como no caso dos modelos Poisson e Binomial, sendo necessário introduzir nestes modelos um parâmetro de dispersão.

Quando o objecto de um determinado estudo ocorre com uma frequência mais elevada e as respostas obtidas apresentam uma variância superior à suportada pelo modelo binomial, estamos perante um fenómeno de sobredispersão (Turkman e Silva 2000). Esta situação pode ficar a dever-se à heterogeneidade entre os indivíduos que não é explicada pelas covariáveis, ou pela existência de uma elevada correlação entre as respostas obtidas levando em algumas situações à formação de clusters na construção de modelos. Na análise de dados ou contagens com características de sobredispersão em que modelos binários não são funcionais são usados modelos como o modelo de quasi-Poisson e o modelo binomial-negativa, que permitem incluir o padrão de correlação entre os indivíduos.

#### 4.1.1. Modelo Poisson, Quasi-Poisson e Binomial-Negativo

O modelo de Poisson sendo um modelo da família exponencial e tendo a particularidade de o seu valor médio ser igual à variância, tem um papel muito importante na análise de dados de contagem. De acordo com Turkman e Silva (2000) ao considerar que as respostas  $Y_i$  são independentes e modeladas por uma distribuição de Poisson de valor médio  $\mu_i$  e que  $\ln(\mu_i) = z_i^T \beta$  com  $i = 1, \dots, n$ , ou seja,

$$f\left(\frac{y_i}{x_i}\right) = \exp\left\{-e^{z_i^T \beta} + y_i^{z_i^T \beta} - \ln y_i!\right\}, \quad y_i = 0, 1, \dots, \quad [4.4]$$

obtendo um modelo linear generalizado com uma função de ligação e conhecido como modelo de regressão de Poisson, e para este modelo a função de ligação é logarítmica.

A utilização do modelo de Poisson impõe a utilização de uma variância igual ao valor médio. No entanto esta situação produz, como referido anteriormente, situações de sobredispersão, que podem ser resolvidas no modelo Poisson com a introdução no modelo de um parâmetro de dispersão  $\phi$ ; contudo, a introdução deste parâmetro no modelo nem sempre conduz a melhores resultados. Outra forma de resolver o problema de sobredispersão é a utilização de modelos como o quasi-Poisson e o modelo binomial-negativo.

No modelo quasi-Poisson assume-se que parâmetro de dispersão  $\phi$ , considera-se que a  $var\left(\frac{y_i}{x_i}\right) = \phi\mu_i$  para  $i = 1, \dots, n$ .

Por outro lado, quando estamos perante dados de contagem com situação de sobredispersão podemos utilizar o modelo binomial negativo, onde se assume uma distribuição binomial negativa para  $\frac{y_i}{x_i}$ , a sua função de probabilidade de densidade (*f. d. p.*) é :

$$f(y, \mu, \theta) = \frac{\Gamma(y+\theta)}{\Gamma(\theta).y!} \cdot \frac{\mu^y.\theta^\theta}{(\mu+\theta)^{y+\theta}} \quad [4.5]$$

onde  $\mu$  é a média,  $\theta$  a forma do parâmetro e  $\Gamma(\cdot)$  a função gamma.

Os métodos como o quasi-poisson, são métodos de “quasi-verosimilhança que permitem a obtenção de estimadores consistentes e assintoticamente normais para  $\beta$ , com apenas uma perda de eficiência. Para que esta perda seja pequena é necessário que a estrutura de variância proposta seja a mais próxima possível da verdadeira estrutura de variância” (Turkman e Silva 2000).

Da mesma forma que o modelo de riscos proporcionais de Cox para a análise de sobrevivência, também os modelos lineares generalizados pressupõem determinado pressupostos apresentados em seguida.

#### 4.1.2. Pressupostos associados aos Modelos lineares generalizados

À semelhança do que acontece nos modelos lineares tradicionais é também fundamental analisar a adequabilidade das suposições associadas aos MLG. Para tal temos que realizar uma análise de diagnóstico ao modelo através dos resíduos e verificando medidas de influência.

Aos modelos lineares generalizados encontra-se associado, como referido anteriormente, um parâmetro de dispersão. Este parâmetro de dispersão pode ser estimado através da máxima verosimilhança. Contudo existe um método mais simples, que se baseia na distribuição de grandes valores de  $n$ , da estatística de Pearson generalizada. Assim, e “assumindo que se aplicou um algoritmo de mínimos quadrados ponderados e se obteve a estimativa de  $\hat{\beta}$  para  $\beta$ , devido à propriedade de invariância dos estimadores de máxima verosimilhança, as estimativas para os parâmetros  $\mu_i$  são dadas por:

$$\hat{\mu}_i = h(z_i^T \hat{\beta}) \quad [4.6]$$

onde a função  $h(\cdot)$  é a inversa da função de ligação” (Turkman e Silva 2000).

Esta estatística de Pearson generalizada, é também útil para avaliar a qualidade de ajustamento de um modelo MLG:

$$X^2 = \sum_i \frac{w_i (y_i - \hat{\mu}_i)^2}{V(\mu_i)} \quad [4.7]$$

No caso dos modelos Poisson e binomial, a estatística  $X^2$ , coincide com a estatística original de Pearson. Para testar a adequabilidade de um modelo utiliza-se  $X^2$  comparando o valor observado com o quantil de probabilidade  $1 - \alpha$  de uma distribuição  $X^2$  com  $n - p$  graus de liberdade, verificando-se contudo que a aproximação pelo  $X^2$  da distribuição de  $X^2$  pode não ser a mais adequada para grandes amostras (Turkman e Silva 2000).

Também os resíduos desviance (desvio residual) padronizados, são utilizados para avaliar a adequabilidade do modelo:

$$R_i^D = \sum_i \frac{\delta_i \sqrt{d_i}}{\sqrt{\hat{\phi}(1-h_{ii})}} \quad [4.8]$$

onde  $\delta_i = \text{sinal}(y_i - \hat{\mu}_i)$  e  $d_i$  é a contribuição da  $i$ -ésima observação para a função desvio.

No que diz respeito à análise da influência esta é verificada recorrendo à estatística de leverage que mede o potencial do ajustamento do modelo ser ou não influenciado pelas covariáveis, avaliando caso a caso.

Relativamente à análise de observações influentes é de destacar uma representação gráfica clássica que ajuda na detecção destas observações, é a da distância de Cook. Esta representação gráfica indica a diferença entre os valores ajustados e os valores sem a observação influente, se esta tivesse sido excluída da análise. Todas as distâncias na representação gráfica devem ter uma magnitude idêntica, quando isto não acontece então temos razões para acreditar que o modelo está a ser enviesado pela presença desta observação.

## 4.2. Resultados

Na aplicação dos modelos lineares generalizados, foi utilizada, como no caso anterior, uma *mega-amostra*, correspondente a todas as mães que tiveram os seus filhos em Portugal nos anos de 2007, 2008 e 2009, com uma dimensão total  $n = 304358$ .

Na elaboração dos modelos MLG devemos ter em consideração a formulação, ajustamento e selecção e validação dos modelos. Na formulação do modelo deveremos ter em consideração a distribuição das variáveis resposta, examinando cuidadosamente os dados; é também importante a escolha e formulação apropriada da matriz tendo em consideração a presença de variáveis dummy. É ainda relevante neste contexto a escolha apropriada da função de ligação, compatível com a distribuição proposta para os dados, não devendo ser uma escolha meramente matemática. Já a fase do ajustamento do modelo implica a estimação dos parâmetros dos modelos, ou seja a estimação dos coeficientes  $\beta$  associados às covariáveis, bem como do parâmetro de dispersão  $\phi$  quando se verifica a sua presença. Por fim, na fase de selecção e validação dos modelos, o objectivo é o de encontrar modelos com um número moderado de parâmetros adequados aos dados em estudo e verificar também a existência de outliers ou observações influentes. Deverão ainda ser ponderados factores como a adequabilidade,

parcimónia e a interpretação do modelo, sendo um bom modelo aquele que consegue um equilíbrio entre estes factores (Turkman e Silva 2000).

Nas últimas décadas em Portugal e por todos os países desenvolvidos assistiu-se à diminuição considerável do número de filhos. Este comportamento tem sido explicado na demografia como estando associado a factores-chave, como o nível de educação da mãe e a sua actividade profissional. Assim, e com o objectivo de compreender o que influencia o número de filhos para as mulheres casadas em Portugal, as variáveis utilizadas foram:

- Duração do casamento,
- Ano de nascimento,
- Idade da Mãe,
- Categoria profissional da Mãe,
- Instrução da Mãe,
- Condição perante o trabalho da Mãe.

Procedeu-se assim à elaboração de três modelos diferentes, o modelo Poisson, o modelo quasi-Poisson e o modelo binomial negativo

#### **4.2.1. Modelo de Poisson e Modelo Quasi-Poisson**

Na tabela 13 apresentam-se os valores  $p$  e AIC para cada um dos modelos estatísticos, relembrando uma vez mais que o modelo Poisson é adequado para contagem mas que nem sempre é o mais adequado quando se verifica situações de sobredispersão. Neste caso, a situação de sobredispersão acontece porque existe um elevado agrupamento de dados/respostas em cada covariável. Ao observar a tabela podemos verificar que todas as variáveis se assumem como altamente significativas para qualquer um dos modelos; no entanto, a covariável “duração do casamento” não é significativa para o modelo Poisson com o parâmetro de dispersão e para o modelo quasi-Poisson.

Relativamente ao valor de *AIC* verifica-se que para o modelo Poisson e para o modelo Poisson com parâmetro de dispersão o valor é igual e que para o modelo quasi-Poisson não obtemos o valor de *AIC* uma vez que este é obtido através da quasi-verosimilhança e não a verosimilhança.

**Tabela 13:** Valores *p* e *AIC* para Modelos Univariados de Poisson

	Modelo Poisson	Modelo Poisson com parâmetro de dispersão	Modelo Quasi-Poisson
(Intercept)	0,00	0,00	0,00
factor(dur)1	0,00	0,21	0,21
factor(ano)2008	0,00	0,66	0,66
factor(ano)20 9	0,00	0,01	0,01
factor(idade)1	0,00	0,02	0,02
factor(prof)1	0,00	0,00	0,00
factor(prof)2	0,02	0,88	0,88
factor(inst)1	0,00	0,00	0,00
factor(inst)2	0,00	0,00	0,00
factor(inst)3	0,00	0,00	0,00
factor(cond)1	0,00	0,00	0,00
Valor AIC	113669	113669	-

Verificou-se na análise anterior que a covariável “duração do casamento” não se apresentava significativa para o modelo, retirando-se esta covariável dos modelos procedemos a nova avaliação dos respectivos valores *p* e *AIC* . Os resultados apresentados (tabela 14) não mostram diferenças significativas nos valores *p* ou nos *AIC* .

**Tabela 14:** Valores  $p$  e AIC para os Modelos Poisson

	Modelo Poisson	Modelo Poisson com parâmetro de dispersão	Modelo Quasi-Poisson
(Intercept)	0,00	0,00	0,00
factor(ano)2008	0,00	0,66	0,66
factor(ano)2009	0,00	0,01	0,01
factor(idade)1	0,00	0,02	0,02
actor(prof)1	0,00	0,00	0,00
factor(prof)2	0,00	0,84	0,84
factor(inst)1	0,00	0,00	0,00
factor(inst)2	0,00	0,00	0,00
factor(inst)3	0,00	0,00	0,00
factor(cond)1	0,00	0,00	0,00
Valor AIC	114051	114051	-

Uma vez mais se deve referir a presença de dados com sobredispersão que “influenciam” os resultados apresentados, devendo assim proceder-se a outra abordagem.

#### 4.2.2. Modelo Binomial-Negativo

A alternativa que nos pareceu mais adequada pela questão da sobredispersão foi o modelo binomial negativo (Faraway 2006). Na tabela 15 apresentamos três modelos onde se encontram os valores  $p$  para cada uma das variáveis; no modelo I verificamos que as covariáveis “duração do casamento” e “ano do nascimento” não são significativas, sendo removida a covariável “ano do nascimento” na elaboração do segundo modelo. Assim no modelo II, verifica-se que a covariável “duração do casamento” volta a apresentar-se como uma covariável não significativa. A remoção desta covariável conduz-nos até ao modelo III onde se verifica que a covariável “idade da mãe” que nos anteriores modelos surgia com duas das categorias não significativas continua com o mesmo comportamento; assim, no modelo IV encontramos as covariáveis significativas: a categoria profissional, o nível de instrução da mãe e a condição perante o trabalho.

No entanto, e ao observar os valores de *AIC* verificamos que o modelo com o valor menor é o modelo I que inclui todas as covariáveis; neste sentido, o modelo final sem interações é o modelo I.

**Tabela 15:** Valores *p* e *AIC* o Modelo Binomial Negativo

	Modelo I	Modelo II	Modelo III	Modelo IV
(Intercept)	0,00	0,00	0,00	0,00
factor(dur)3-5	0,11	0,11	-	-
factor(dur)6-8	0,28	0,28	-	-
factor(dur)9-30	0,34	0,37	-	-
factor(ano)2008	0,82	-	-	-
factor(ano)2009	0,14	-	-	-
factor(idade)30-33	0,58	0,53	0,58	-
factor(idade)34-36	0,00	0,00	0,00	-
factor(idade)37-51	0,25	0,19	0,23	-
factor(prof)1	0,31	0,00	0,22	0,26
factor(prof)2	0,00	0,00	0,00	0,00
factor(inst)1	0,00	0,00	0,00	0,00
factor(inst)2	0,00	0,00	0,00	0,00
factor(inst)3	0,00	0,00	0,00	0,00
factor(cond)1	0,00	0,00	0,00	0,00
Valor AIC	8966	8967	8967	8990

Escolhido o modelo final sem interações procedeu-se à construção de modelos com interações. Através do teste de razão de verosimilhanças concluiu-se que o modelo mais adequado acabou por ser o modelo sem interações. Na tabela 16 encontramos a informação relativa ao modelo final onde encontramos os valores das estimativas  $\beta$ , os valores exponenciais das estimativas, o desvio padrão, o valor de *t* e o valor *p*.

No modelo final (o modelo I) verifica-se assim que todas as covariáveis apresentadas apresentam uma correlação com a evolução do número de nascimentos em Portugal. Ao observar os valores das categorias da covariável “duração do casamento” verificamos que relativamente à categoria de referência (duração do casamento inferior a 3 anos), são as mulheres com duração de casamento superior aos 9 anos que apresentam possibilidades mais baixas de ter filhos. Por outro lado, no que diz respeito ao ano do nascimento verifica-se, relativamente a 2007, que no ano de 2008 a possibilidade de ter

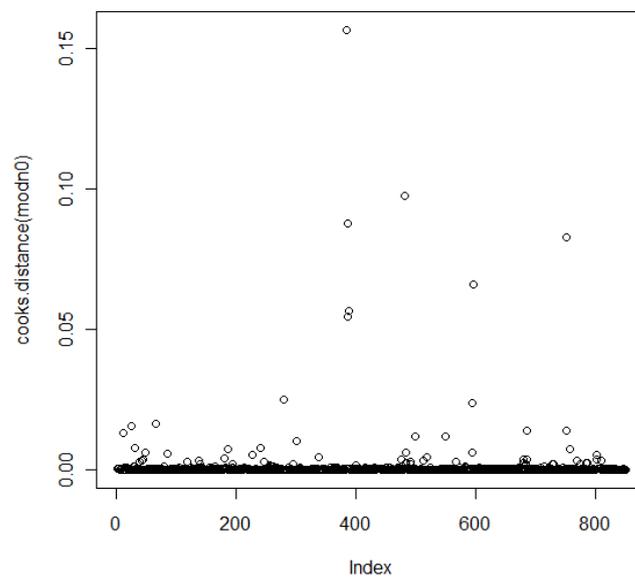
um filho aumentou 2,9% e em 2009 17%. A covariável “idade da mãe” apresenta na categoria 34-36 anos a possibilidade mais elevada de ter filhos em comparação com a categoria de referência (16-29), com um valor de aproximadamente 40%. Valor também elevado se encontra na relação entre as categorias profissionais onde se verifica que as mães na categoria profissional 2 apresentam 12% de possibilidades de ter filhos e a categoria 3 uma possibilidade de 47%, isto quando comparadas com a categoria de referência 1. Finalmente, e no que diz respeito à covariável “instrução da mãe” verifica-se que as mães com menores possibilidades de terem filhos são as mães com o nível de instrução 3 (ensino superior) com uma possibilidade de 3 vezes mais que as mães da categoria de referencia (sem instrução), e são as mães com o nível de instrução 1 (ensino básico) as com maiores probabilidades de terem filhos (13 vezes mais).

Tabela 16: Modelo Binomial Negativo final

	Estimativa ( $\beta$ )	Exponencial da estimativa	Desvio padrão	Valor de $t$	Valor $p$
(Intercept)	4,10	1,26	0,25	1,61	0,00
factor(dur)3-5	0,23	1,17	0,14	1,61	0,11
factor(dur)6-8	0,16	1,15	0,15	1,08	0,28
factor(dur)9-30	0,14	0,97	0,14	0,95	0,34
factor(ano)2008	-0,03	0,83	0,13	-0,23	0,82
factor(ano)2009	-0,19	0,92	0,13	-1,49	0,14
factor(idade)30-33	-0,08	0,60	0,14	-0,56	0,58
factor(idade)34-36	-0,51	0,85	0,14	-3,53	0,00
factor(idade)37-51	-0,17	0,88	0,15	-1,14	0,25
factor(prof)2	-0,13	0,53	0,13	-1,01	0,31
factor(prof)3	-0,64	13,03	0,13	-49,99	0,00
factor(inst)1	2,57	4,12	0,21	12,02	0,00
factor(inst)2	1,42	3,08	0,21	6,63	0,00
factor(inst)3	1,13	0,10	0,21	5,28	0,00
factor(cond)2	-2,27	1,26	0,10	-21,80	0,00

Analisado o modelo é ainda importante, tal como referido anteriormente, verificar a adequabilidade do mesmo e verificar os pressupostos associados a este modelo. Ao analisar a distância de Cook para as observações influentes (figura 64), verificamos que a presença de algumas observações influentes; no entanto, valores entre -2 e 2 não são preocupantes ou considerados observações severas.

**Figura 64:** Distância de Cook



Ao observar os resíduos leverage (figura 65a) verificamos que, para a questão da análise da variância das observações, embora se destaquem algumas observações, estas não apresentam valores “severos” que permitam afirmar uma influência no modelo final. Já no diz respeito à representação gráfica dos resíduos deviance (figura 65b), verifica-se que este modelo é adequado.

**Figura 65:** Resíduos Leverage e Resíduos Desviance

Figura 65a: Resíduos Leverage

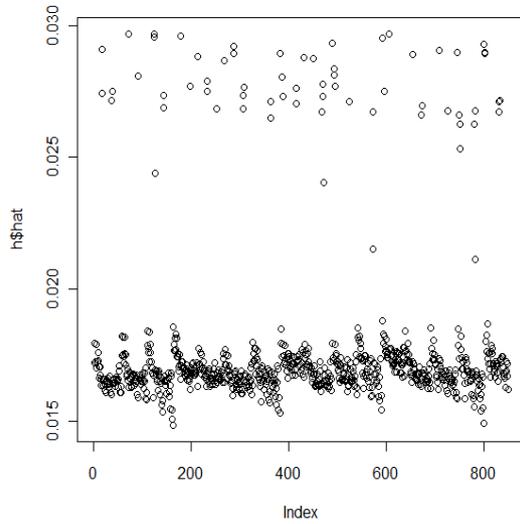
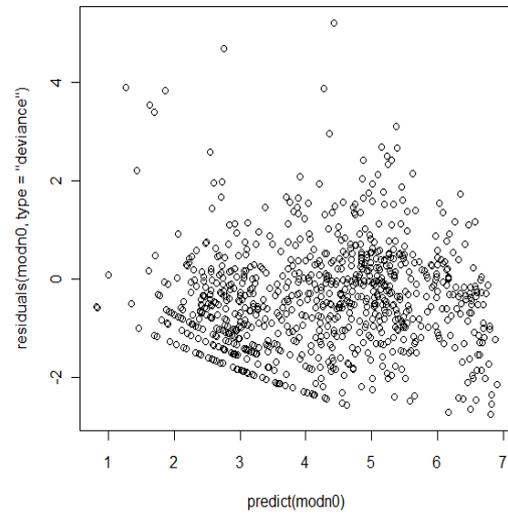
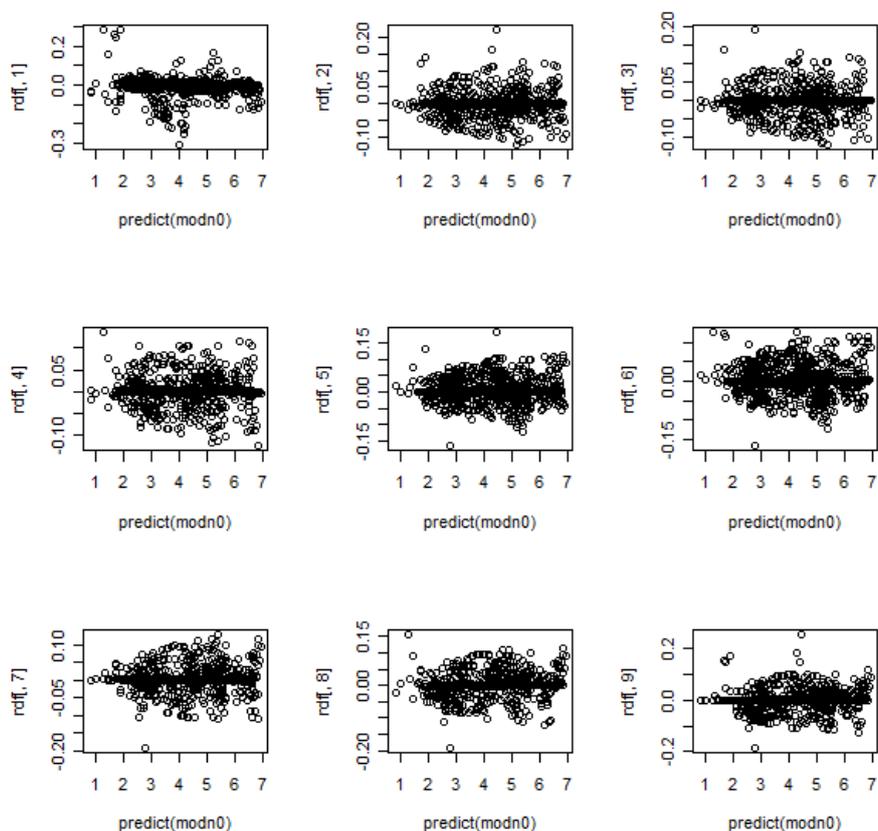


Figura 65b : Resíduos Desviance



Por fim, e no que diz respeito aos resíduos de Pearson padronizados (figura 66) e reforçando que nem sempre esta abordagem é a mais indicada para amostras de tamanho muito grande, observa-se que para cada uma das predições os valores se concentram em torno de zero, não se apresentando totalmente aleatórios.

**Figura 66:** Resíduos de Pearson padronizados



É contudo importante compreender a capacidade explicativa do modelo através de:

$$\widehat{\rho^2} = 1 - \frac{\frac{RD}{gl RD}}{\frac{ND}{gl ND}} \Rightarrow 1 - \frac{\frac{1107,6}{834}}{\frac{2269}{848}} = 0,504$$

O modelo final apresentado, onde se incluíram todas as covariáveis propostas inicialmente e onde se excluem quaisquer interações, explica aproximadamente 50% dos resultados obtidos.

Para a compreensão dos factores que influenciam o número dos nascimentos em Portugal, foram realizados modelos lineares generalizados, onde se verificou que no período de 2007 a 2009, a duração do casamento, o ano de nascimento, a idade da mãe, categoria profissional, a instrução e a condição perante o trabalho da mãe, são factores significativos na explicação deste comportamento. Verificou-se, assim, nesta análise

que as mães em casamentos com mais de 9 anos apresentam uma possibilidade de ter filhos inferior à de mulheres casadas há menos tempo. Verificou-se também que, em comparação com o ano de 2007, no ano de 2008, a possibilidade de ter filhos aumentou aproximadamente 3% e no ano de 2009 aumentou para 17%. Observou-se igualmente que as mães com o ensino básico têm possibilidades bastante mais elevadas que as restantes de terem filhos, sendo que são também as mães com a categoria profissional 3 que apresentam possibilidades mais elevadas para o nascimento de um filho.

## 5. Discussão dos resultados

A diminuição do número de nascimentos, característica acentuada no final do século passado e início deste século, alastrou a toda a Europa, América do Norte e ainda aos países mais desenvolvidos da Ásia, como o Japão, atingindo valores do índice sintético de fecundidade nunca antes alcançados. Portugal não sendo a excepção neste contexto, e embora os seus níveis de fecundidade nunca tenham ficado abaixo de 1,3 filhos por mulher, considerados por muito autores como o valor mínimo limite para uma intervenção social dramática, apresenta uma situação difícil no que diz respeito aos seus níveis de fecundidade.

O acentuado declínio da fecundidade surge-nos na literatura associado muitas vezes a questões como o prolongamento de formação e do ensino, a forte participação feminina no mercado de trabalho, as melhorias acentuadas das condições de vida, o uso alargado de contraceptivos a todas as mulheres (que de acordo com alguns autores levou à diminuição do número de casamentos – Goldin e Katz 2002), entre outros factores, que no seu conjunto, são reconhecidos no contexto da demografia, como a segunda transição demográfica. A segunda transição demográfica apresenta assim como características fundamentais o adiamento da fecundidade, a diminuição acentuada do número de filhos, a diminuição do número de casamentos, o aumento do grau de instrução e o consequente aumento da participação no mercado de trabalho pelas mães.

Neste contexto de declínio de fecundidade, procedeu-se ao longo deste trabalho a uma análise estatística com o objectivo de compreender o impacto destas características no espaçamento decorrido entre o casamento e o nascimento do primeiro filho, entre o nascimento do primeiro e do segundo filho e entre o nascimento do segundo e do terceiro filho, utilizando-se métodos estatísticos no contexto de Análise de Sobrevivência (Modelo de Risco Proporcional de Cox e Modelo Aditivo de Aalen). Por outro lado, e sendo também importante compreender o que influencia o número de filhos das mães portuguesas efectuou-se a análise dos factores que influenciam o número de nascimentos através da utilização de Modelos Lineares Generalizados (Modelo Poisson, Modelo Quasi-Poisson e Modelo Binomial Negativo).

No espaçamento decorrido entre o casamento e o nascimento do primeiro filho, no período entre 1995 e 2007, observou-se que as covariáveis que influenciaram o espaçamento e a decisão de ter filhos foram a condição perante o trabalho, a idade da

mãe e a categoria profissional. Verificou-se assim que as mães desempregadas esperam menos tempo após o seu casamento para o nascimento do primeiro filho, o mesmo acontece a mães mais jovens e a mães na categoria profissional 1 (onde se enquadram quadros superiores da administração pública, dirigentes, quadros superiores de empresas ou especialistas das profissões intelectuais e científicas). Por outro lado, são as mães com mais idade que esperam mais tempo após o casamento para o nascimento do primeiro filho e em situação idêntica encontramos as mães na categoria profissional 3 (onde se enquadram mulheres agricultoras, operárias ou trabalhadoras não qualificadas). São assim as mulheres com situações profissionais menos favoráveis que por mais tempo adiam a fecundidade ao primeiro filho dentro do casamento e aquelas que possuem idades mais avançadas.

Na evolução do espaçamento decorrido entre o nascimento do primeiro e do segundo filho, para o período entre 1995 a 2005, identificaram-se as covariáveis filiação do nascimento, a instrução, a categoria profissional e a idade da mãe, como influenciadoras do espaçamento. No espaçamento entre o nascimento do primeiro e do segundo filho, foram as mães fora do casamento as que apresentaram espaçamentos mais elevados entre os nascimentos. No contexto do nível de instrução da mãe observou-se que as mães com o ensino superior esperam menos tempo até ao nascimento do segundo filho. A covariável categoria profissional das mães permitiu-nos compreender que as mães da categoria profissional 3 apresentam um espaçamento mais elevado entre o nascimento do primeiro e do segundo filho. No grupo de idades foi possível observar que no início do período (1995) eram as mães mais novas que registavam um espaçamento mais elevado entre os dois nascimentos, situação que se inverteu passando no ano de 2005 a ser característica do grupo de idades mais elevadas. Conclui-se assim que o espaçamento entre o primeiro e segundo filhos é maior para mulheres fora de um casamento, assim como para as mães com categoria profissional 3, e que ainda que o comportamento no que diz respeito ao grupo de idade da mãe sofreu uma alteração deslocando um maior espaçamento para idades mais avançadas.

Ainda no período entre 1995 e 2005, foi analisada a evolução no espaçamento entre o nascimento do segundo e terceiro filhos, para o qual se identificaram as covariáveis, filiação do nascimento, a instrução, a condição perante o trabalho, a categoria profissional e a idade da mãe, como factores influenciadores do adiamento. Após o nascimento do segundo filho, e no que diz respeito à covariável filiação, nos primeiros

cinco anos não existem diferenças entre nascimentos ocorridos dentro ou fora do casamento; no entanto, após este período observou-se que são as mães fora do casamento que apresentam espaçamento entre os nascimentos maior. Por outro lado, são as mães ao nível do ensino superior as que apresentam um espaçamento menor entre o nascimento do segundo e terceiro filhos. Já na condição perante o trabalho, observa-se que não existem diferenças entre mães em situação de emprego ou desemprego. Na covariável categoria profissional observou-se que as mães da categoria 1 apresentam espaçamentos menores até ao nascimento do seu terceiro filho, e as mães da categoria 3 apresentam espaçamentos mais. Nas idades das mães observou-se uma proximidade de comportamentos entre as mães mais novas e as mais velhas sendo que estas são as que apresentam espaçamentos mais elevados entre o nascimento do segundo e terceiro filhos. No espaçamento entre o segundo e o terceiro filho identificou-se, assim, que nos primeiros cinco anos após o nascimento do segundo filho são semelhantes os comportamentos para mães casadas ou não, e também que são as mães com o ensino superior que apresentam espaçamentos menores entre os nascimentos.

Na construção dos modelos de riscos proporcionais de Cox, não foram obtidos resultados significativos para afirmar a proporcionalidade dos modelos. Contudo, os resultados apresentados ao longo de todo trabalho, bem como a análise do índice de prognóstico são reveladores de que os modelos finais vão de encontro aos factores empíricos explicativos dos espaçamentos entre o casamento e o nascimento do primeiro filho, entre o nascimento do primeiro e segundo filhos e, por fim, entre o nascimento do segundo e terceiro filhos.

Na observação dos factores que potencialmente influenciam o número de nascimentos no período entre 2007 e 2009, verificou-se que as mães em casamentos com mais de 9 anos apresentam a possibilidade de ter filhos inferior à de mulheres casadas há menos tempo. Verificou-se também que, em comparação com o ano de 2007, no ano de 2008 a possibilidade de ter filhos aumentou aproximadamente 3% e no ano de 2009 aumentou para 17%. Observou-se que as mães com o ensino básico têm possibilidades bastante mais elevadas que as restantes de terem filhos, sendo que são também as mães com a categoria profissional 3 as que apresentam possibilidades mais elevadas relativamente ao nascimento de um filho. É, no entanto, importante referir que o modelo apresentado através da metodologia de modelos lineares generalizados explica apenas cerca de 50% do comportamento.

A abordagem estatística utilizada ao longo deste trabalho possibilitou a análise e compreensão mais pormenorizadas dos comportamentos associados à fecundidade num período recente. No entanto, do ponto de vista unicamente estatístico os resultados aqui apresentados não são significativos; todavia, devemos ter em conta que estamos a estudar uma “mega amostra” ou população respeitante a todas as mulheres que tiveram filhos em Portugal no período entre 1995 e 2009, num total de 1648682 observações.

Considerando que o trabalho incidiu sobre o estudo de humanos, a ciência estatística deve no futuro colaborar com a ciência demográfica que aborda estas questões com perspectivas mais adequadas ao comportamento demográfico como é a fecundidade. Questões como pressupostos associados aos modelos, perdem o seu importante significado da estatística quando utilizados no contexto demográfico, por muitas vezes não se encontrarem disponíveis amostras mas apenas populações.

Podemos assim concluir que embora do ponto de vista estatístico os resultados obtidos não sejam totalmente significativos, do ponto de vista demográfico os modelos apresentados reflectem os resultados empíricos de extrema relevância para a compreensão do comportamento recente da fecundidade em Portugal e possível estimação do seu futuro.

## 6. Bibliografia

**Andersson, G.** (2005), “A study on policies and practices in the selected countries that encourage childbirth: The case of Sweden”, in *MPIDR working paper*, WP 2005-005.

**Andersson, G., & Scott, K.**, (2005). “Labour-market status and first-time parenthood: the experience of immigrant women in Sweden, 1981–97”, *Population Studies* 59(1): 21–38.

**Billari, F.** (2004): “Becoming an Adult in Europe: a Macro (/Micro)- Demographic Perspective”, *Demographic Research*, Special Collection 3, Article 2, April 2004, pp. 15-44, Max Planck Institute for Demographic Research, Rostock, [www.demographic-research.org](http://www.demographic-research.org).

**Bongaarts, J.** (1999): “Fertility Decline in the Developed World: where will it end?” *The American Economic Review*, May, 89, 2, pp. 256-260.

**Bongaarts, J. & G. Feeney**, (1998), “On the quantum and tempo of fertility”, in *Population and Development Review*, 24: 271-291.

**Brewster, K. & Rindfuss, R.** (2000), “Fertility and Women’s employment in industrialized nations”, in *Annual Review Sociology*: 26, pp. 271-296.

**Colosimo & Giolo** (2006), *Análise de Sobrevivência Aplicada*, Editora Edgar Blucher.

**Caleiro, A.; Mendes, M.; Rego, C.**, (2006). “Educação e Fecundidade em Portugal: As diferenças nos níveis de educação influenciam as taxas de fecundidade?”, *Documento de Trabalho n° 2006/10*, Universidade de Évora, Departamento de Economia.

**Cunha, V.**, (2004) “ A fecundidade das Famílias Portuguesas”, in Wall, K. (coord.), *Famílias no Portugal Contemporâneo*, Lisboa, Imprensa de Ciências Sociais/ICS.

**Esping-Andersen, G.** (1999). *Social Foundations of Postindustrial Economies*. Oxford: Oxford University Press.

**Faraway, J.J.** (2006), *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*, Taylor & Francis.

**Frejka, T. & Sardon, J. P.**, (2006). “First birth trends in developed countries: Persisting parenthood postponement”, *Demographic Research*, Volume 15, Article 6, September

2006, pp. 147-180, Max Planck Institute for Demographic Research, Rostock, [www.demographicresearch.org](http://www.demographicresearch.org).

**Frejka, T. & Sobotka, T.** (2008); “Fertility in Europe: Diverse, delayed and below replacement”, Volume 19, Article 3, pp 15-46, Max Planck Institute for Demographic Research, Rostock, <http://www.demographic-research.org/Volumes/Vol19/3/>

**Grant, J., Hoorens, Sivadsan, Davanzo, Hale, Gibson, & Butz** (2004). *Low Fertility in Population Aging: Causes, Consequences and Policy Options*. Santa Monica, C.A.: RAND, [www.rand.org](http://www.rand.org).

**Goldin, C. & Katz, L. F.** (2002). “The power of the pill: Oral contraceptives and women’s career and marriage decisions”. in [\*Journal of Political Economy\* 110: 730-770](#)

**Hoem, J.** (2005), “ Why does Sweden have such high fertility?”, in *Demographic Research*, Volume 13, Article 22, pp. 559-572, Rostock

**Hosmer D.W., Lemeshow S.** (1989). *Applied Logistic Regression*. New York: Wiley.

**Hosmer D.W., Lemeshow S.**( 1999). *Applied Survival Analysis*. New York: Wiley.

**Kohler, H. & Ortega, J. A.,** (2002). “Tempo-adjusted period parity progression measures: assessing the implications of delayed childbearing for cohort fertility in Sweden, the Netherlands and Spain”, *Demographic Research*, Volume 6, Article 7, Mar 2002, pp. 145-190, Max Planck Institute for Demographic Research, Rostock [www.demographic-research.org](http://www.demographic-research.org).

**Kohler, H., Billari, F. & Ortega, J. A.,** (2002). “The emergency of lowest-low fertility in Europe during the 1990s”, *Population and Development Review*, 28 (4), pp. 641-680.

**Kohler, H., Billari, F. & Ortega, J. A.,** (2006), “Low and Lowest-Low Fertility in Europe: Causes, Implications and Policy options”, in F. R. Harris (Ed.), *The Baby Bust: Who will do the Work? Who Will Pay the Taxes?* Lanham, MD: Rowman & Littlefield Publishers, 48-109.

**Lesthaeghe, R. & Neels, K.** (2002). “From the first to the second demographic transition: An interpretation of the spatial continuity of demographic innovation in France, Belgium and Switzerland”. *European Journal of Population* 18 (4): 325-360.

**Lesthaeghe, R. & Surkyn., J.** (1988). “Cultural dynamics and economic theories of fertility change.” in *Population and Development Review* 14 (1): 1-45.

**Maroco, J.** (2007). *Análise estatística com utilização do SPSS*. Edições Sílabo, Lisboa.

**Maroco, J. & Bispo, R.** (2005). *Estatística aplicada às ciências sociais e humanas*. Climepsi Editores, Lisboa.

**Matsuo, H.** 2003. *The transition to motherhood in Japan. A comparison with the Netherlands*. Doctoral thesis, University of Groningen. Population Studies, Rozenberg Publishers, Amsterdam.

**McCullagh, P. & J. A. Nelder** (1989). *Generalized Linear Model* (2nd ed.). Monographs on Statistics Applied Probability. London: Chapman & Hall.

**Mendes, M.,** (1992) “Análise sociodemográfica do declínio da fecundidade da população portuguesa na década de 90”, *Universidade de Évora*, policopiado.

**Mendes, M., Guerreiro, G., & Caleiro, A.,**(2005): “Fertility in Portugal, How persistent is it?”, Working Paper nº2005/12, Universidade de Évora, Departamento de Economia.

**Mendes, F., & Rego, C.** (2007): “Baixa fecundidade nos países do Sul da Europa: a importância das desigualdades na educação e na participação no mercado de trabalho, ao nível regional”, in 13º Congresso da APDR, Açores, Portugal.

**Neyer, G.** (2003), “Family Policies and low fertility in western Europe”, in *MPDID Working Paper WP 2003-021*, Max Planck Institute, Rostock.

**Rocha, C., & Papoila, A. L.** (2009). *Análise de Sobrevivência*. Publicação disponibilizada in XVII Congresso da Sociedade Portuguesa de Estatística. Sociedade Portuguesa de Estatística.

**Rodriguez, G.** (2009). Lecture Notes for WWS509. Princeton University. Available at <http://data.princeton.edu/wws509/notes/>.

**Siegel, J. S. & Swanson, D.A.** (2004). *The methods and materials of demography*, Elsevier Academic Press.

**Silva, F.**(2002), “Projecção das taxas de fecundidade específicas por idades no horizonte de longo prazo (2001-2050): Estudo de um modelo de previsão com séries temporais”, in *Revista de Estudos Demográficos* n°32, INE, Lisboa.

**Sobotka T.**, (2003): “Tempo-Quantum and Period-Cohort Interplay in Fertility Changes in Europe. Evidence from the Czech Republic, Italy, the Netherlands and Sweden”, *Demographic Research*, Volume 8, Article 6, April 2003, pp. 151-214, Max Planck Institute for Demographic Research, Rostock, [www.demographic-research.org](http://www.demographic-research.org);

**Sobotka T.**, (2004), “Is lowest-low fertility in Europe explained by the postponement of childbearing?”, in *Population and Development Review*, 30: 195-220.

**Turkman, M.A.M, & Silva, G.L.**, (2000) Modelos Lineares Generalizados da teoria à prática,, Edições SPE, Lisboa

**van de Kaa, D. J.**, (2002).“The Idea of a Second Demographic Transition in Industrialized Countries.”, in *Sixth Welfare Policy Seminar*, Tokyo: National Institute of Population and Social security.

**Vikat, A.** (2004), “Women's Labour Force Attachment and Childbearing in Finland”, in *Demographic Research, Special Collection 3*, Article 8, Rostock.

**Yavuz, S.** (2008). Fertility Decline in Turkey from the 1980s Onwards: Patterns by Main Language Groups. Ankara: HUIPS [Doctoral Thesis].

## Anexo A

### Scripts R Análise de Sobrevida

#### A. 1- Espaçamento entre o casamento e o nascimento do primeiro filho

##### **Exemplo do ano de 1995**

```
dados23<-read.table("CasPrimFilh95.txt", header=T)
attach (dados23)
dados23
require (survival)

### Avaliar as covariáveis individualmente ####
###Instrução da mãe recodificada###
ekmisnrmae_categ<-survfit(Surv(tempo,censura)~isnrmae_categ)
summary (ekmisnrmae_categ)
Surv (tempo, censura)
kmisnrmae_categ<-survfit(Surv(tempo, censura)~isnrmae_categ)
kmisnrmae_categ
plot(kmisnrmae_categ, main="Instrução da mãe", col= 1:4)
legend(15, 0.99, col=c(1:4), lty=1:1, c("Sem instrução", "Ensino Básico", "Ensino Secundário", "Ensino Superior"), bty="n", cex=0.9)
survdif(Surv(tempo, censura==1)~isnrmae_categ, rho=1)#####Peto

###Condição perante o trabalho###
ekmcdtbmae<-survfit(Surv(tempo,censura)~cdtbmae)
summary (ekmcdtbmae)
Surv (tempo, censura)
kmcdbmae<-survfit(Surv(tempo, censura)~cdtbmae)
kmcdbmae
plot(kmcdbmae, main="Condição perante o trabalho", col= 1:2)
legend(15, 0.99, col=c(1:2), lty=1:1, c("Empregado", "Desempregado"), bty="n", cex=0.9)
survdif(Surv(tempo, censura==1)~cdtbmae, rho=0)#####Log-Rank

###Profissão da mãe recodificada###
ekmprofmae_categ<-survfit(Surv(tempo,censura)~profmae_categ)
summary (ekmprofmae_categ)
Surv (tempo, censura)
kmprofmae_categ<-survfit(Surv(tempo, censura)~profmae_categ)
kmprofmae_categ
plot(kmprofmae_categ, main="Categoria Profissional (Profissão)", col= 1:3)
legend(15, 0.99, col=c(1:3), lty=1:1, c("Categoria 1", "Categoria 2", "Categoria 3"), bty="n", cex=0.9)
survdif(Surv(tempo, censura==1)~profmae_categ, rho=1)#####Peto

###Situação Profissional da mãe###
ekmsitpfmae<-survfit(Surv(tempo,censura)~sitpfmae)
summary (ekmsitpfmae)
Surv (tempo, censura)
```

```

kmsitpfmae<-survfit(Surv(tempo, censura)~sitpfmae)
kmsitpfmae
plot(kmsitpfmae, main="Situação Profissional", col= 1:4)
legend(10, 0.99, col=c(1:4), lty=1:1, c("Empregador", "Trabalhador por conta própria",
"Trabalhador por conta de outrém", "Outra situação"), bty="n", cex=0.9)
survdifff(Surv(tempo, censura==1)~sitpfmae, rho=1)#####Peto

####Idade da Mãe categorizada ano a ano ####
ekmidademaecateg<-survfit(Surv(tempo,censura)~idademaecateg)
summary (ekmidademaecateg)
Surv (tempo, censura)
kmidademaecateg<-survfit(Surv(tempo, censura)~idademaecateg)
kmidademaecateg
plot(kmidademaecateg, main="Grupo de idade da mãe", col= 1:13)
survdifff(Surv(tempo, censura==1)~idademaecateg, rho=1)#####Peto

####Idade da Mãe categorizada quantil global ####
ekmidademaecateg2<-survfit(Surv(tempo,censura)~idademaecateg2)
summary (ekmidademaecateg2)
Surv (tempo, censura)
kmidademaecateg2<-survfit(Surv(tempo, censura)~idademaecateg2)
kmidademaecateg2
plot(kmidademaecateg2, main="Grupo de idade da mãe", col= 1:4)
legend(20, 0.99, col=c(1:4), lty=1:1, c("16-25", "26-27", "28-30","31-54"), bty="n",
cex=0.9)
survdifff(Surv(tempo, censura==1)~idademaecateg2, rho=1)#####Peto

####Ano do casamento categorizado em décadas####
ekmanocasa_catg<-survfit(Surv(tempo,censura)~anocasa_catg)
summary (ekmanocasa_catg)
Surv (tempo, censura)
kmanocasa_catg<-survfit(Surv(tempo, censura)~anocasa_catg)
kmanocasa_catg
plot(kmanocasa_catg, main="Ano do casamento por décadas", col= 1:3)
legend(17, 0.99, col=c(1:3), lty=1:1, c("Década 70", "Década 80", "Década 90"),
bty="n", cex=0.9)
survdifff(Surv(tempo, censura==1)~anocasa_catg, rho=0)#####Log rank

####Modelo ajustado com as variáveis significativas####
fit1<-coxph(Surv(tempo,
censura)~factor(isntrmaecateg)+factor(cdtbmae)+factor(profmaecateg)+factor(sitpfmae)
+factor(idademaecateg)+factor(anocasa_catg), data=dados23, method="breslow",
x=TRUE)
fit1
cox.zph(fit1)

```

```

####Retirar a variável menos significativa no fit 1####
fit2<-coxph(Surv(tempo,
censura)~factor(isnrmae_categ)+factor(cdtbmae)+factor(profmae_categ)+factor(sitpfm
ae)+factor(idadema_e_categ), data=dados23, method="breslow", x=TRUE)
fit2
cox.zph(fit2)

```

```

####Retirar a segunda variável menos significativa no fit 2####
fit3<-coxph(Surv(tempo,
censura)~factor(isnrmae_categ)+factor(cdtbmae)+factor(profmae_categ)+factor(idade
mae_categ), data=dados23, method="breslow", x=TRUE)
fit3
cox.zph(fit3)

```

```

####Acrescentar interações####
fit4<-coxph(Surv(tempo,
censura)~factor(isnrmae_categ)+factor(cdtbmae)+factor(profmae_categ)+factor(idade
mae_categ)+factor(isnrmae_categ)*factor(cdtbmae)+factor(isnrmae_categ)*factor(pro
fmae_categ)+factor(isnrmae_categ)*factor(idadema_e_categ)+factor(cdtbmae)*factor(p
rofmae_categ)+factor(cdtbmae)*factor(idadema_e_categ)+factor(profmae_categ)*factor(
idadema_e_categ), data=dados23, method="breslow", x=TRUE)
fit4
cox.zph(fit4)

```

```

####Retirar interações não significativas####
fit5<-coxph(Surv(tempo,
censura)~factor(isnrmae_categ)+factor(cdtbmae)+factor(profmae_categ)+factor(idade
mae_categ)+factor(isnrmae_categ)*factor(cdtbmae)+factor(isnrmae_categ)*factor(pro
fmae_categ)+factor(cdtbmae)*factor(profmae_categ)+factor(cdtbmae)*factor(idadema_e
_categ)+factor(profmae_categ)*factor(idadema_e_categ),data=dados23,method="breslo
w", x=TRUE)
fit5
cox.zph(fit5)

```

```

fit6<-coxph(Surv(tempo,
censura)~factor(isnrmae_categ)+factor(cdtbmae)+factor(profmae_categ)+factor(idade
mae_categ)+factor(isnrmae_categ)*factor(cdtbmae)+factor(cdtbmae)*factor(profmae_
categ)+factor(cdtbmae)*factor(idadema_e_categ)+factor(profmae_categ)*factor(idadema
e_categ), data=dados23, method="breslow", x=TRUE)
fit6
cox.zph(fit6)

```

```

fit7<-coxph(Surv(tempo,
censura)~factor(isnrmae_categ)+factor(cdtbmae)+factor(profmae_categ)+factor(idade
mae_categ)+factor(cdtbmae)*factor(profmae_categ)+factor(cdtbmae)*factor(idadema_e
_categ)+factor(profmae_categ)*factor(idadema_e_categ),data=dados23,
method="breslow", x=TRUE)
fit7
cox.zph(fit7)

```

```

fit8<-coxph(Surv(tempo,
censura)~factor(isntrmae_categ)+factor(cdtbmae)+factor(profmae_categ)+factor(idade
mae_categ)+factor(cdtbmae)*factor(idadema_e_categ)+factor(profmae_categ)*factor(ida
dema_e_categ), data=dados23, method="breslow", x=TRUE)
fit8
cox.zph(fit8)

```

```

fit9<-coxph(Surv(tempo,
censura)~factor(isntrmae_categ)+factor(cdtbmae)+factor(profmae_categ)+factor(idade
mae_categ)+factor(cdtbmae)*factor(idadema_e_categ),data=dados23,
method="breslow", x=TRUE)
fit9
cox.zph(fit9)

```

```

fit10<-coxph(Surv(tempo,
censura)~factor(isntrmae_categ)+factor(cdtbmae)+factor(profmae_categ)+factor(idade
mae_categ), data=dados23, method="breslow", x=TRUE)
fit10
cox.zph(fit10)
rs<-cox.zph(fit10)

```

### Verificação do Pressuposto de Riscos Proporcionais ###

```

# Teste de Harrell's
resid(fit10,type="scaledsch") #Resíduos de Schoenfeld
zph<-cox.zph(fit10, transform="identity") ## g(t) = t
zph

```

###Instrução da mãe#####

```

par(mfrow=c(1,3))
plot(rs[1])
plot(rs[2])
plot(rs[3])
###Condição perante o trabalho#####
plot(rs[4])
par(mfrow=c(1,2))#####Profissão da mãe###
plot(rs[5])
plot(rs[6])
par(mfrow=c(1,3))###Idade da mãe#####
plot(rs[7])
plot(rs[8])
plot(rs[9])

```

####Deviance & Martingala####

```

par(mfrow=c(1,2))
rd<-resid(fit10,type="deviance") # resíduos deviance
rm<-resid(fit10,type="martingale") # resíduos martingala
p1<-fit10$linear.predictors
plot(p1,rm, xlab="Preditor linear", ylab="Resíduo martingal", pch=16)
plot(p1,rd, xlab="Preditor linear", ylab="Resíduo deviance" , pch=16)

```

```

####Influência#####
infl<-apply(dfbetas,1,sum)
plot(1:fit10$n,infl, xlab="Índice",ylab="Influência")
abline(h=0)

####Dfbetas###
par(mfrow=c(1,2))
dfbetas<-resid(fit10,type="dfbeta")
plot(factor(isnrmae_categ),dfbetas[,1], xlab="isnrmae_categ", ylab="Nível de
Instrução")
plot(factor(cdtbmae), dfbetas[,2], xlab="cdtbmae", ylab="Condição perante o
trabalho")
plot(factor(profmae_categ), dfbetas[,3], xlab="profmae_categ", ylab="Categoria
Profissional (profissão)")
plot(factor(idadema_e_categ),dfbetas[,4], xlab="idadema_e_categ", ylab="Grupo de
idade da mãe")

source("Rfun.r")
plot.pi (fit10, main="Índice de Prognóstico")

```

#### A.2- Espaçamento entre o nascimento do primeiroe segundo filho

##### **Exemplo do ano de 1995**

```

dados1<-read.table("PrimSegFilh95.txt", header=T)
attach (dados1)
dados1
require (survival)

####Avaliar as variáveis####
####Filiação###
ekmfil<-survfit(Surv(tdanunv,censura)~fil)
summary (ekmfil)
Surv (tdanunv, censura)
kmfil<-survfit(Surv(tdanunv, censura)~fil)
kmfil
plot(kmfil, main="Filiação", col= 1:2)
legend(15, 0.99, col=c(1:2), lty=1:1, c("Dentro do Casamento", "Fora do Casamento"),
bty="n", cex=0.9)
survdifff(Surv(tdanunv, censura==1)~fil, rho=1)#####Peto

####Instrução da mãe recodificada###
ekminstrmae_categ<-survfit(Surv(tdanunv,censura)~instrmae_categ)
summary (ekminstrmae_categ)
Surv (tdanunv, censura)
kminstrmae_categ<-survfit(Surv(tdanunv, censura)~instrmae_categ)
kminstrmae_categ
plot(kminstrmae_categ, main="Instrução da mãe", col= 1:4)

```

```

legend(15, 0.99, col=c(1:4), lty=1:1, c("Sem instrução", "Ensino Básico", "Ensino
Secundário", "Ensino Superior"), bty="n", cex=0.9)
survdifff(Surv(tdanunv, censura==1)~instrmae_categ, rho=1)#####Peto

```

```

###Condição perante o trabalho###

```

```

ekmcdtbmae<-survfit(Surv(tdanunv,censura)~cdtbmae)
summary (ekmcdtbmae)
Surv (tdanunv, censura)
kmcdbmae<-survfit(Surv(tdanunv, censura)~cdtbmae)
kmcdbmae
plot(kmcdbmae, main="Condição perante o trabalho", col= 1:2)
legend(15, 0.99, col=c(1:2), lty=1:1, c("Empregado", "Desempregado"), bty="n",
cex=0.9)
survdifff(Surv(tdanunv, censura==1)~cdtbmae, rho=1)#####Peto

```

```

###Profissão da mãe recodificada###

```

```

ekmprofmae_categ<-survfit(Surv(tdanunv,censura)~profmae_categ)
summary (ekmprofmae_categ)
Surv (tdanunv, censura)
kmprofmae_categ<-survfit(Surv(tdanunv, censura)~profmae_categ)
kmprofmae_categ
plot(kmprofmae_categ, main="Categoria Profissional (Profissão)", col= 1:3)
legend(15, 0.99, col=c(1:3), lty=1:1, c("Categoria 1", "Categoria 2", "Categoria 3"),
bty="n", cex=0.9)
survdifff(Surv(tdanunv, censura==1)~profmae_categ, rho=1)#####Peto

```

```

###Situação Profissional da mãe###

```

```

ekmsitpfmae<-survfit(Surv(tdanunv,censura)~sitpfmae)
summary (ekmsitpfmae)
Surv (tdanunv, censura)
kmsitpfmae<-survfit(Surv(tdanunv, censura)~sitpfmae)
kmsitpfmae
plot(kmsitpfmae, main="Situação Profissional", col= 1:4)
legend(10, 0.99, col=c(1:4), lty=1:1, c("Empregador", "Trabalhador por conta própria",
"Trabalhador por conta de outrém", "Outra situação"), bty="n", cex=0.9)
survdifff(Surv(tdanunv, censura==1)~sitpfmae, rho=1)#####Peto

```

```

###Idade da Mãe ###

```

```

ekmidademaecateg2<-survfit(Surv(tdanunv,censura)~idademaecateg2)
summary (ekmidademaecateg2)
Surv (tdanunv, censura)
kmidademaecateg2<-survfit(Surv(tdanunv, censura)~idademaecateg2)
kmidademaecateg2
plot(kmidademaecateg2, main="Grupo de idade da mãe", col= 1:4)
legend(20, 0.99, col=c(1:4), lty=1:1, c("13-25", "26-28", "29-32", "33-54"), bty="n",
cex=0.9)
survdifff(Surv(tdanunv, censura==1)~idademaecateg2, rho=1)#####Peto

```

```

####Modelo ajustado com as variáveis significativas####
fit1<-coxph(Surv(tdanunv,
censura)~factor(fil)+factor(instrmae_categ)+factor(cdtbmae)+factor(profmae_categ)+fa
ctor(sitpfmae)+factor(idadema_categ2), data=dados1, method="breslow", x=TRUE)
fit1
cox.zph(fit1)

fit2<-coxph(Surv(tdanunv,
censura)~factor(fil)+factor(instrmae_categ)+factor(profmae_categ)+factor(sitpfmae)+fa
ctor(idadema_categ2), data=dados1, method="breslow", x=TRUE)
fit2
cox.zph(fit2)

fit3<-coxph(Surv(tdanunv,
censura)~factor(fil)+factor(instrmae_categ)+factor(profmae_categ)+factor(idadema_c
ateg2), data=dados1, method="breslow", x=TRUE)
fit3
cox.zph(fit3)

####Interacções####
fit4<-coxph(Surv(tdanunv,
censura)~factor(fil)+factor(instrmae_categ)+factor(profmae_categ)+factor(idadema_c
ateg2) + factor(fil)*factor(instrmae_categ)+ factor(fil)*factor(profmae_categ)
+factor(fil)*factor(idadema_categ2)+factor(instrmae_categ)*factor(profmae_categ)
+factor(instrmae_categ)*factor(idadema_categ2)+factor(profmae_categ)*factor(idade
mae_categ2), data=dados1, method="breslow", x=TRUE)
fit4
cox.zph(fit4)

####Retirar as interacções####
fit5<-coxph(Surv(tdanunv,
censura)~factor(fil)+factor(instrmae_categ)+factor(profmae_categ)+factor(idadema_c
ateg2)+ factor(fil)*factor(instrmae_categ)+
factor(fil)*factor(profmae_categ)+factor(fil)*factor(idadema_categ2)
+factor(instrmae_categ)*factor(idadema_categ2)+factor(profmae_categ)*
factor(idadmae_categ2), data=dados1, method="breslow", x=TRUE)
fit5 cox.zph(fit5)

anova (fit3, fit5, test="Chisq")

cox.zph(fit5)
rs<-cox.zph(fit5)

par(mfrow=c(1,2))####Filiação####
plot(rs[1])

par(mfrow=c(1,3))####Instrução da mãe####
plot(rs[2])

```

```

plot(rs[3])
plot(rs[4])

par(mfrow=c(1,2))####Profissão da mãe###
plot(rs[6])

par(mfrow=c(1,2))###Idade da mãe####
plot(rs[7])
plot(rs[8])

####Deviance & Martingala####
par(mfrow=c(1,2))
rd<-resid(fit5,type="deviance") # resíduos deviance
rm<-resid(fit5,type="martingale") # resíduos martingala
p1<-fit5$linear.predictors
plot(p1,rm, xlab="Preditor linear", ylab="Resíduo martingal", pch=16)
plot(p1,rd, xlab="Preditor linear", ylab="Resíduo deviance" , pch=16)

###Dfbetas###
par(mfrow=c(1,1))
dfbetas<-resid(fit5,type="dfbeta")
plot(factor(fil),dfbetas[,1], xlab="fil", ylab="Filiação")
plot(factor(instrmae_categ),dfbetas[,2], xlab="instrmae_categ", ylab="Nível de
Instrução")
plot(factor(cdtbmae), dfbetas[,3], xlab="cdtbmae", ylab="Condição perante o
trabalho")
plot(factor(profmae_categ), dfbetas[,4], xlab="profmae_categ", ylab="Categoria
Profissional (profissão)")
plot(factor(idademaie_categ2),dfbetas[,5], xlab="idademaie_categ2", ylab="Grupo de
idade da mãe")

###Influência#####
infl<-apply(dfbetas,1,sum)
plot(1:fit5$n,infl, xlab="Índice",ylab="Influência")
abline(h=0)

source("Rfun.r")
plot.pi (fit5, main="Índice de Prognóstico")

```

### A.3 - Espaçamento entre o nascimento do segundo e terceiro filho

#### **Exemplo do ano de 1995**

```

dados1<-read.table("SegTerFilh95.txt", header=T)
attach (dados1)
dados1
require (survival)

```

```

###Avaliar as variáveis####
###Filiação###
ekmfil<-survfit(Surv(tdanunv,censura)~fil)
summary (ekmfil)
Surv (tdanunv, censura)
kmfil<-survfit(Surv(tdanunv, censura)~fil)
kmfil
plot(kmfil, main="Filiação", col= 1:2)
legend(15, 0.99, col=c(1:2), lty=1:1, c("Dentro do Casamento", "Fora do Casamento"),
bty="n", cex=0.9)
survdiff(Surv(tdanunv, censura==1)~fil, rho=1)#####Peto

###Instrução da mãe recodificada###
ekminstrmae_categ<-survfit(Surv(tdanunv,censura)~instrmae_categ)
summary (ekminstrmae_categ)
Surv (tdanunv, censura)
kminstrmae_categ<-survfit(Surv(tdanunv, censura)~instrmae_categ)
kminstrmae_categ
plot(kminstrmae_categ, main="Instrução da mãe", col= 1:4)
legend(15, 0.99, col=c(1:4), lty=1:1, c("Sem instrução", "Ensino Básico", "Ensino
Secundário", "Ensino Superior"), bty="n", cex=0.9)
survdiff(Surv(tdanunv, censura==1)~instrmae_categ, rho=1)#####Peto

###Condição perante o trabalho###
ekmcdtbmae<-survfit(Surv(tdanunv,censura)~cdtbmae)
summary (ekmcdtbmae)
Surv (tdanunv, censura)
kmcdbmae<-survfit(Surv(tdanunv, censura)~cdtbmae)
kmcdbmae
plot(kmcdbmae, main="Condição perante o trabalho", col= 1:2)
legend(15, 0.99, col=c(1:2), lty=1:1, c("Empregado", "Desempregado"), bty="n",
cex=0.9)
survdiff(Surv(tdanunv, censura==1)~cdtbmae, rho=1)#####Peto

###Profissão da mãe recodificada###
ekmprofmae_categ<-survfit(Surv(tdanunv,censura)~profmae_categ)
summary (ekmprofmae_categ)
Surv (tdanunv, censura)
kmprofmae_categ<-survfit(Surv(tdanunv, censura)~profmae_categ)
kmprofmae_categ
plot(kmprofmae_categ, main="Categoria Profissional (Profissão)", col= 1:3)
legend(15, 0.99, col=c(1:3), lty=1:1, c("Categoria 1", "Categoria 2", "Categoria 3"),
bty="n", cex=0.9)
survdiff(Surv(tdanunv, censura==1)~profmae_categ, rho=1)#####Peto

###Situação Profissional da mãe###
ekmsitpfmae<-survfit(Surv(tdanunv,censura)~sitpfmae)
summary (ekmsitpfmae)

```

```

Surv (tdanunv, censura)
kmsitpfmae<-survfit(Surv(tdanunv, censura)~sitpfmae)
kmsitpfmae
plot(kmsitpfmae, main="Situação Profissional", col= 1:4)
legend(10, 0.99, col=c(1:4), lty=1:1, c("Empregador", "Trabalhador por conta própria",
"Trabalhador por conta de outrém", "Outra situação"), bty="n", cex=0.9)
survdiff(Surv(tdanunv, censura==1)~sitpfmae, rho=1)#####Peto

###Idade da Mãe categorizada###
ekmidademaecateg2<-survfit(Surv(tdanunv,censura)~idademaecateg2)
summary(ekmidademaecateg2)
Surv (tdanunv, censura)
kmidademaecateg2<-survfit(Surv(tdanunv, censura)~idademaecateg2)
kmidademaecateg2
plot(kmidademaecateg2, main="Grupo de idade da mãe", col= 1:3)
legend(20, 0.99, col=c(1:3), lty=1:1, c("16-33", "34-36","37-52"), bty="n", cex=0.9)
survdiff(Surv(tdanunv, censura==1)~idademaecateg2, rho=1)#####Peto

###Modelo ajustado com as variáveis significativas###
fit1<-coxph(Surv(tdanunv,
censura)~factor(fil)+factor(instrmaecateg)+factor(cdtbmae)+factor(profmaecateg)+fa
ctor(idademaecateg2), data=dados1, method="breslow", x=TRUE)
fit1

fit2<-coxph(Surv(tdanunv,
censura)~factor(fil)+factor(instrmaecateg)+factor(cdtbmae)+factor(idademaecateg2),
data=dados1, method="breslow", x=TRUE)
fit2

###Resíduos deviance(outliers) e martingala###
par(mfrow=c(1,2))
rd<-resid(fit1,type="deviance") # resíduos deviance
rm<-resid(fit1,type="martingale") # resíduos martingala
p1<-fit1$linear.predictors
plot(p1,rm, xlab="Preditor linear", ylab="Resíduo martingal", pch=16)
plot(rm)
abline(h=0, col=2, lty=2)
plot(age, rm, xlab = "Idade", ylab = "Resíduo")
lines(lowess(age, rm, iter = 0), lty = 2)
plot(p1,rd, xlab="Preditor linear", ylab="Resíduo deviance" , pch=16)

###Resíduos score (observações influentes)###
par(mfrow=c(2,3))
dfbetas<-resid(fit1,type="dfbeta") #resíduos score
plot(fil,dfbetas[,1], xlab="fil", ylab="Filiação")
plot(factor(instrmaecateg), dfbetas[,2], xlab="instrmaecateg", ylab="Nível de
instrução da mãe")
plot(factor(fil), dfbetas[,3], xlab="", ylab="")
plot(factor(),dfbetas[,4], xlab="", ylab="")

```

```

###Gráficos de resíduos dfbetas associados a cada covariável vs os valores da
respectiva covariável###
infl<-apply(dfbetas,1,sum)
plot(1:fit1$N,infl, xlab="Índice",ylab="Influência")
abline(h=0)

plot(1:fit1$N, dfbetas[, 3], xlab="Índice", ylab="Resíduos")
plot(1:fit1$N, dfbetas[, 4], xlab="Índice", ylab="Resíduos")

#####Índice de prognóstico#####
source("Rfun.r")
plot.pi (fit1, main="Índice de Prognóstico")

###Interacções###
fit3<-coxph(Surv(tdanunv,
censura)~factor(fil)+factor(instrmae_categ)+factor(cdtbmae)+factor(idadema_
e_categ2)
+factor(fil)*factor(instrmae_categ)+factor(fil)*factor(cdtbmae)+factor(fil)*factor(idade
mae_categ2)+factor(instrmae_categ)*factor(cdtbmae)+factor(instrmae_categ)*factor(id
adema_
e_categ2)+factor(cdtbmae)*factor(idadema_
e_categ2),data=dados1,method="bresl
ow", x=TRUE)
fit3
cox.zph(fit3)

###Retirar as interacções###
fit4<-coxph(Surv(tdanunv,
censura)~factor(fil)+factor(instrmae_categ)+factor(cdtbmae)+factor(idadema_
e_categ2)
+factor(fil)*factor(instrmae_categ)+factor(fil)*factor(cdtbmae)+factor(fil)*factor(idade
mae_categ2)+factor(instrmae_categ)*factor(idadema_
e_categ2)+factor(cdtbmae)*factor(
idadema_
e_categ2), data=dados1, method="breslow", x=TRUE)
fit4
cox.zph(fit4)

fit5<-coxph(Surv(tdanunv,
censura)~factor(fil)+factor(instrmae_categ)+factor(cdtbmae)+factor(idadema_
e_categ2)
+factor(fil)*factor(instrmae_categ)+factor(fil)*factor(cdtbmae)+factor(instrmae_categ)
*factor(idadema_
e_categ2)+factor(cdtbmae)*factor(idadema_
e_categ2),data=dados1, meth
od="breslow", x=TRUE)
fit5
cox.zph(fit5)

anova (fit2, fit5, test="Chisq")

cox.zph(fit5)
rs<-cox.zph(fit5)
par(mfrow=c(1,3)) ###É só mudar ou n usar....###
plot(rs)###Todos###

par(mfrow=c(1,2))###Filiação#####
plot(rs[1])

```

```

par(mfrow=c(1,3))###Instrução da mãe####
plot(rs[2])
plot(rs[3])
plot(rs[4])

par(mfrow=c(1,2))###Condição####
plot(rs[5])

par(mfrow=c(1,2))###Idade da mãe####
plot(rs[6])
plot(rs[7])

####Deviance & Martingala####
par(mfrow=c(1,2))
rd<-resid(fit5,type="deviance") # resíduos deviance
rm<-resid(fit5,type="martingale") # resíduos martingala
p1<-fit5$linear.predictors
plot(p1,rm, xlab="Preditor linear", ylab="Resíduo martingal", pch=16)
plot(p1,rd, xlab="Preditor linear", ylab="Resíduo deviance" , pch=16)

###Dfbetas###
par(mfrow=c(1,1))
dfbetas<-resid(fit5,type="dfbeta")
plot(factor(fil),dfbetas[,1], xlab="fil", ylab="Filiação")
plot(factor(instrmae_categ),dfbetas[,2],xlab="instrmae_categ",ylab="Nível de
Instrução")
plot(factor(cdtbmae), dfbetas[,3], xlab="cdtbmae", ylab="Condição perante o trabalho")
plot(factor(profmae_categ),dfbetas[,4],xlab="profmae_categ",ylab="Categoria
Profissional (profissão)")
plot(factor(idadema_e_categ2),dfbetas[,5],xlab="idadema_e_categ2",ylab="Grupo de
idade da mãe")

####Influência#####
infl<-apply(dfbetas,1,sum)
plot(1:fit5$n,infl, xlab="Índice",ylab="Influência")
abline(h=0)

source("Rfun.r")
plot.pi (fit5, main="Indice de Prognóstico")

```

## **Anexo B**

### **Scripts R Modelos Lineares Generalizados**

#### *B.1 – Modelo Poisson e Modelo quasi-Poisson*

```
dados78<-read.table("glmnascIII.txt", header=T)
attach(dados78)
dados78
dur<-factor(dur)
ano<-factor(ano)
idade<-factor(idade)
prof<-factor(prof)
inst<-factor(inst)
cond<-factor(cond)
modp78<-glm(nasc~factor(dur)+factor(ano)+factor(idade)+factor(prof)+factor(inst)
+factor(cond),family=poisson, dados78)
summary(modp78)

modp1 <- glm(nasc ~ factor(dur),family=poisson, dados78)
summary(modp1)

modp2 <- glm(nasc ~ factor(dur)+factor(ano),family=poisson, dados78)
summary(modp2)

modp3 <- glm(nasc ~ factor(dur)+factor(ano)+factor(idade),family=poisson, dados78)
summary(modp3)

modp4 <- glm(nasc ~ factor(dur)+factor(ano)+factor(idade),family=poisson, dados78)
summary(modp4)

require(epicalc)
lrtest(modp3, modp1)

# Modelo com parâmetro de dispersão #
pd78 <-sum(residuals(modp78,type="pearson")^2/modp78$df.res)
pd78
summary (modp78, dispersion=pd78)
drop1(modp,test="F")

# Alternativa - quasipoisson#
modqp78<-glm(nasc~factor(dur)+factor(ano)+factor(idade)+factor(prof)+factor(inst)
+factor(cond),family=quasipoisson, dados78)
summary (modqp78)

###Sem a variável duração###
ano<-factor(ano)
```

```

idade<-factor(idade)
prof<-factor(prof)
inst<-factor(inst)
cond<-factor(cond)
modp79<-glm(nasc~+factor(ano)+factor(idade)+factor(prof)+factor(inst)
+factor(cond),family=poisson, dados78)
summary(modp79)

```

```

# Modelo com parâmetro de dispersão #
pd79 <-sum(residuals(modp79,type="pearson")^2/modp79$df.res)
pd79
summary (modp79, dispersion=pd79)
drop1(modp79,test="F")

```

```

modqp79<-glm(nasc~factor(ano)+factor(idade)+factor(prof)+factor(inst)
+factor(cond),family=quasipoisson, dados78)
summary (modqp79)

```

## B.2 – Modelo Binomial Negativo

```

dados5<-read.table("glmnasc.txt", header=T)
attach(dados5)
dados5
library(MASS)
dur<-factor(dur)
dur
ano<-factor(ano)
ano
idade<-factor(idade)
idade
prof<-factor(prof)
inst<-factor(inst)
cond<-factor(cond)
cond

```

```

modn0 <- glm(nasc ~
factor(dur)+factor(ano)+factor(idade)+factor(prof)+factor(inst)+factor(cond) ,
family=negative.binomial(1),dados5, maxit= 100, trace=TRUE)
modn0
summary(modn0)

```

```

modn1 <- glm(nasc ~ factor(dur)+factor(idade)+factor(prof)+factor(inst)+factor(cond) ,
family=negative.binomial(1),dados5, maxit= 100, trace=TRUE)
modn1
summary(modn1)

```

```

modn2 <- glm(nasc ~ factor(idade)+factor(prof)+factor(inst)+factor(cond) ,
family=negative.binomial(1),dados5, maxit= 100, trace=TRUE)
modn2
summary(modn2)

```

```

####Interacções####

```

```

modn3 <- glm(nasc ~
factor(idade)+factor(prof)+factor(inst)+factor(cond)+factor(idade)*factor(prof)+factor(i
dade)*factor(inst)+factor(idade)*factor(cond)
+factor(prof)*factor(inst)+factor(prof)*factor(cond)+factor(inst)*factor(cond),
family=negative.binomial(1),dados5, maxit= 100, trace=TRUE)
modn3
summary(modn3)
exp(modn3$coef)

```

```

modn4 <- glm(nasc ~
factor(idade)+factor(prof)+factor(inst)+factor(cond)+factor(idade)*factor(prof),
family=negative.binomial(1),dados5, maxit= 100, trace=TRUE)
modn4
summary(modn4)
exp(modn4$coef)

```

```

modn5 <- glm(nasc ~
factor(idade)+factor(prof)+factor(inst)+factor(cond)+factor(idade)*factor(prof)+factor(i
dade)*factor(inst), family=negative.binomial(1),dados5, maxit= 100, trace=TRUE)
modn5
summary(modn5)
exp(modn5$coef)

```

```

modn6 <- glm(nasc ~
factor(idade)+factor(prof)+factor(inst)+factor(cond)+factor(idade)*factor(prof)+factor(i
dade)*factor(inst)+factor(idade)*factor(cond), family=negative.binomial(1),dados5,
maxit= 100, trace=TRUE)
modn6
summary(modn6)
exp(modn6$coef)

```

```

1-pchisq(modn0$deviance - modn6$deviance, modn0$df.residual - modn6$df.residual)

```

```

anova (modn2, modn3 , test="Chisq")

```

```

###Análise de residuos####

```

```

plot (predict(modn0), residuals (modn0, type="deviance"))
sum(residuals(modn0,type="pearson")^2)

```

```

rdf<-dfbetas(modn0)
par (mfrow=c(3,2))
plot(predict(modn0), rdf[,1])
plot(predict(modn0), rdf[,2])
plot(predict(modn0), rdf[,3])
plot(predict(modn0), rdf[,4])
plot(predict(modn0), rdf[,5])
plot(predict(modn0), rdf[,6])
plot(predict(modn0), rdf[,7])
plot(predict(modn0), rdf[,8])
plot(predict(modn0), rdf[,9])
plot(predict(modn0), rdf[,10])
plot(predict(modn0), rdf[,11])
plot(predict(modn0), rdf[,12])
plot(predict(modn0), rdf[,13])
plot(predict(modn0), rdf[,14])
plot(predict(modn0), rdf[,15])
plot(predict(modn0), rdf[,16])

library (car)
library (faraway)
halfnorm(residuals(modn0)) #####manter um padrão#####
plot(cooks.distance(modn0)) ###Pontos influentes###
h <- influence(modn0)
plot(h$hat) #####Levarege , repercursão#####
halfnorm(h$hat)
d<-halfnorm(h$hat)

```