



UMA CARACTERIZAÇÃO DOS IMIGRANTES NO DISTRITO DE ÉVORA

RELATÓRIO DE ESTÁGIO

MESTRADO EM MODELAÇÃO ESTATÍSTICA E ANÁLISE DE DADOS

ZULMIRA XIMENES DA COSTA

185041

Sob orientação:

Professor Doutor Paulo Infante

Professora Doutora Maria Filomena Mendes

Inspector João Carlos Agostinho

ÉVORA

2011

Agradecimentos

À Fundação das Universidades Portugueses (FUP) em Timor Leste que me ofereceu a oportunidade de trabalhar no âmbito do curso de informática em Timor Leste.

À Universidade Nacional Timor Loro Sa'e (UNTL) que me indicou para receber uma bolsa fornecida pelo Governo Português através de IPAD.

Ao Instituto Português de Apoio de Desenvolvimento (IPAD) que financiou a minha vinda e a minha estadia em Portugal, tornando possível que eu realiza-se o curso de mestrado.

Ao professor Paulo Infante como orientador científico que durante o curso me ajudou e explicou as matérias de estatística e resolveu todos os problemas que enfrentei nas aulas do curso. Com o seu esforço e a sua dedicação consegui realizar este curso até ao final.

À professora Filomena Mendes como co-orientadora do relatório de estágio e também ao inspector João Agostinho como orientador Profissional do estágio.

Aos professores da Universidade de Évora, especialmente do Departamento de Matemática no mestrado de Modelação Estatística e Análise de Dados.

Ao professor José Carlos Tiago de Oliveira como o meu tutor da bolsa para o meu curso de mestrado.

Ao Serviço de Estrangeiros e Fronteiras (SEF) do Distrito de Évora que me facilitou realizar o estágio e o acesso os dados sobre os imigrantes a partir de 2006 até 2009, agradeço o acolhimento, a paciência, a oportunidade e os conhecimentos que me foram oferecidos durante a minha participação neste estágio.

Aos funcionários do SEF (João Lopes, Maria da Conceição Courelas, Maria Manuela Baptista, Joaquim Guerra e Maria Nazaré Madeira) que me facilitaram e transmitiram todas as actividades do SEF durante o processo de estágio.

À mana Vanda Narciso que me ajudou dando coragem durante o estudo e também me ajudou corrigir a minha língua portuguesa.

A todos os que comentaram, criticaram, sugeriram, ajudaram e apoiaram este meu trabalho.

À Dália e ao Pedro Nogueira, pela sua amizade e pela disponibilidade que sempre tiveram para ajudar me.

À Mana Susana, pela sua amizade e pela disponibilidade que sempre tiveram para ajudar me.

À Mana Maria José, Francisca e Joaquina pela vossa amizade e convívio durante a minha estadia em Évora.

A minha família, pela disponibilidade que sempre apresentaram para me ajudar no que fosse necessário. Pelo apoio incondicional que sempre me deram mesmo que ao longe da distância.

Ao Valente Ferreira que durante toda a estadia me acompanhou e ajudou sempre que necessário.

A todos os amigos de mestrado na Universidade de Évora.

Finalmente, agradeço a todos aqueles que me ajudaram, nas mais variadas formas, na elaboração deste trabalho.

Em especial, venho manifestar o meu orientador, Professor Paulo Infante, a minha profunda gratidão por toda a ajuda, amizade e compreensão que foram indispensáveis na concretização deste trabalho e durante o caminho do meu mestrado.

Chegou a hora da partida...não é um momento fácil mas um dia iria acontecer! Obrigada a todos os colegas e amigos que durante todos estes anos fizeram parte da minha vida! Para os meus Amigos, obrigada pelo convívio, pelo apoio e principalmente pela amizade.

Espero voltar a encontrar-vos em Portugal... Em língua tétum HAU HADOMI TIMOR LESTE E PORTUGAL!!!

Zulmira Ximenes da Costa

ÍNDICE

Agradecimentos	i
Índice	iii
Índice das Figuras	vi
Índice dos Quadros	ix
Índice das Tabelas	xi
Resumo	xiii
Abstract	xiv
Capítulo 1 Introdução	
1.1. Preliminares	1
1.2. Contextualização do Problema	1
1.3. Actividades dos Serviços de Estrangeiros e Fronteiras	4
1.3.1. Sector documental	4
1.3.2. Sector operacional	5
1.4. Actividades do Estágio no SEF	6
Capítulo 2 Tópicos Gerais e Metodologia	
2.1. A Imigração	8
2.1.1 Conceito	8
2.1.2 Razões da imigração	9
2.1.3 Política da imigração	10
2.1.4 Casamentos dos imigrantes	12
2.2. Metodologia do Trabalho	13
2.2.1. Local e tempo de estágio	13
2.2.2. População e amostra	13
2.2.3. Variáveis de investigação	14
2.2.4. Técnica de recolher os dados	14
2.2.5. Técnica de análise dos dados	14
2.3. Análise Exploratória de Dados	15
2.4. Análise de Sobrevivência	16
2.4.1. O estimador de Kaplan-Meier	18
2.4.2. Modelo Semi-paramétrico de Cox	23
2.4.3. Modelo probabilísticos análise de sobrevivência	31

2.4.4. Estimaco dos parmetros do modelo	35
2.4.5. Mtodos dos grficos para escolher o melhor modelo	35
2.5. Modelos Fragilidade	39
Captulo 3 Caracterizao dos Imigrantes do Distrito de vora	
3.1 Descrio das Variveis	40
3.2 Anlise Exploratrio de Dados	40
3.3 Associao de Variveis	62
3.3.1. Relao entre sexo e pas de origem	62
3.3.2. Relao entre o de ano entrada em vora e a nacionalidade	64
3.3.3. Relao entre designao da morada e pases Unio Europeia	69
3.3.4. Sexo e designao da morada	70
3.3.5. Relao entre nacionalidade e sexo	71
3.3.6. Relao entre estado civil e o sexo	72
3.3.7. Relao entre nvel escolaridade e o sexo	74
3.3.8. Relao entre sexo e motivo de entrada	75
3.3.9. Relao entre pases de origem e motivo de entrada	77
3.3.10. Relao entre pas origem e profisso dos imigrantes	78
3.4. Teste t Para Amostras Emparelhadas	80
Captulo 4 Estudo Estatstico do Casamento dos Imigrantes	
4.1. Modelao no paramtrica	83
4.1.1. Curvas de Kaplan-Meier	84
4.1.2. Modelo de Regresso de Cox	97
4.1.3. Modelos Paramtricos	107
4.1.4. Modelos de Fragilidade	111
Captulo 5 Concluses e Recomendaes	
5.1. Concluses	112
5.2. Recomendaes	114
BIBLIOGRAFIA	116

Índice dos Anexos

Anexo 1. Descrição das variáveis	119
Anexo 2. Descrição das covariáveis para análise sobrevivência	125
Anexo 3. O Script do software R para Análise sobrevivência	126

Índice das Figuras

Figura 1. Gráfico circular por sexo	41
Figura 2. Diagrama de barras por nacionalidade dos imigrantes	43
Figura 3. Diagrama barras por codificação da nacionalidade	44
Figura 4. Diagramas barras por populações imigrantes Europeus de 2006 a 2009	44
Figura 5. Diagramas barras por populações imigrantes Brasileiros de 2006 a 2009	44
Figura 6. Diagramas barras por populações imigrantes Africanos de 2006 a 2009	44
Figura 7. Diagramas barras por populações imigrantes Restantes de 2006 a 2009	44
Figura 8. Gráfico circular por nível de escolaridade	46
Figura 9. Histograma e curva normal para idade dos imigrantes	47
Figura 10. Caixa de bigode por idade	47
Figura 11. Diagrama de barra por estado civil	48
Figura 12. Diagrama de barra por motivo de entrada	49
Figura 13. Diagrama de circular por profissão	50
Figura 14. Diagrama de circular por tipo de celebração do casamento	51
Figura 15. Diagrama de circular por União Europeia	52
Figura 16. Diagrama barras por naturalidade dos imigrantes	54
Figura 17. Diagrama barras por codificação da naturalidade dos imigrantes	55
Figura 18. Diagrama de barra para os imigrantes que mantêm nacionalidade e adquiriram outra nacionalidade	55
Figura 19. Diagrama de barra para naturalidade dos cônjuges	57
Figura 20. Diagrama de barra para a nacionalidade dos cônjuges	58
Figura 21. Diagrama de circular por tipo profissão dos cônjuges	59
Figura 22. O histograma para idade dos cônjuges	59
Figura 23. Caixa de bigode para idade dos cônjuges	60
Figura 24. Diagrama de barra para números dos filhos do casamento	61
Figura 25. Diagrama de barra por números de filhos de outro casamento	62
Figura 26. Diagrama de barra em relação ano de entrada e nacionalidade codificados por classes	65
Figura 27. Diagrama de barra em relação ano de entrada e nacionalidade codificada por classes (Juntando União europeia pelo SEF e Município)	67
Figura 28. Curva do estimador Kaplan-Meier da função de sobrevivência para sexo	86

Figura 29. Curva do estimador Kaplan-Meier da função de sobrevivência para tipo celebração do casamento	87
Figura 30. Curva do estimador Kaplan-Meier da função de sobrevivência para país de origem	87
Figura 31. Curva do estimador Kaplan-Meier da função de sobrevivência para motivo de entrada	89
Figura 32. Curva do estimador Kaplan-Meier da função de sobrevivência para nível escolaridade	91
Figura 33. Curva do estimador Kaplan-Meier da função de sobrevivência para nível escolaridade	93
Figura 34. Curva do estimador Kaplan-Meier da função de sobrevivência para profissão dos imigrantes	94
Figura 35. Curva do estimador Kaplan-Meier da função de sobrevivência para profissão dos cônjuges	94
Figura 36. Curva do estimador Kaplan-Meier da função de sobrevivência para nacionalidade dos imigrantes	95
Figura 37. Curvas do estimador Kaplan-Meier da função de sobrevivência para idade dos imigrantes codificados por duas e por quatro classes	96
Figura 38. Curvas de sobrevivência para existência dos filhos do casamento e de outro casamento	97
Figura 39. $\text{Log}(\hat{\Lambda}_{0j}(t))$ versus tempo para as covariáveis	102
Figura 40. Resíduos de padronizados de Schoenfeld versus tempos para as covariáveis seleccionados	103
Figura 41. Resíduos martingal e deviance versus predictor linear do modelo de cox final ajustados para os dados de casamento	104
Figura 42. Resíduos martingal	104
Figura 43. Resíduos dos pontos influentes	105
Figura 44. Gráficos de t versus $-\log(\hat{S}(t))$, $\log(t)$ versus $\log(\hat{S}(t))$ e $\log(t)$ versus $\Phi^{-1}(\hat{S}(t))$	108
Figura 45. Gráficos das sobrevivências estimadas por Kaplan-Meier versus as sobrevivências estimadas pelos modelos Log-normal e Log-logístico	108
Figura 46. Gráficos das sobrevivências estimadas por Kaplan-Meier versus as sobrevivências estimadas pelos modelos Exponencial e Weibull	109

Figura 47. Curvas de sobrevivências estimadas pelos modelos de Log-normal e Log-logístico	110
Figura 48. Curvas de sobrevivências estimadas pelos modelos de Exponencial e Weibull	110

Índice dos Quadros

Quadro 1. Relação entre o sexo e países de origem	63
Quadro 1. Estatística teste do Qui-quadrado	63
Quadro 2. O risco estimado	64
Quadro 3. Quadro ano de entrada em Évora e Nacionalidade	64
Quadro 4. Estatística teste para o Qui-quadrado	65
Quadro 6. Relação entre país de origem e ano de entrada	68
Quadro 7. Estatística teste do Qui-quadrado	68
Quadro 8. Tabela de contingência entre designação da morada e países da União Europeia	69
Quadro 9. Estatística teste do Qui-quadrado	69
Quadro 10. Tabela de contingência entre sexo e designação da morada	70
Quadro 11. Estatística teste do Qui-quadrado	70
Quadro 12. Tabela de contingência da codificação da nacionalidade e o sexo	71
Quadro 13. Estatística teste do Qui-quadrado	71
Quadro 14. Tabela de contingência do sexo e estado civil	72
Quadro 15. Tabela de contingência do sexo e codificação do estado civil	73
Quadro 16. Estatística teste do Qui-quadrado	73
Quadro 17. Tabela de contingência do sexo e do nível de escolaridade	74
Quadro 18. Estatística teste do Qui-quadrado	74
Quadro 19. Tabela de contingência do sexo e do motivo de entrada	75
Quadro 20. Tabela de contingência do sexo e codificação do motivo de entrada	76
Quadro 21. Estatística teste do Qui-quadrado	76
Quadro 22. Tabela de contingência do país da União Europeia e do motivo da entrada	77
Quadro 23. Tabela de contingência do país da União Europeia e da codificação do motivo de entrada	78
Quadro 24. Estatística teste do Qui-quadrado	78
Quadro 25. Tabela de contingência do país da União Europeia e da profissão dos imigrantes	79
Quadro 26. Tabela de contingência do país da União Europeia e da codificação da profissão dos imigrantes	79
Quadro 27. Estatística teste do Qui-quadrado	80
Quadro 28. Teste a normalidade	81

Quadro 29. Ranks	82
Quadro 30. Estatística teste	82
Quadro 31. Tempo médio e mediana da sobrevivência para a covariável país de origem (País de União Europeia)	88
Quadro 32. Percentis para a covariável país de origem (País de União Europeia)	88
Quadro 33. Tempo médio e mediana da sobrevivência para a covariável motivo de entrada	90
Quadro 34. Percentis para a covariável motivo de entrada	90
Quadro 35. Tempo médio e mediana da sobrevivência para a covariável nível escolaridade	92
Quadro 36. Percentis para a covariável motivo de entrada	92

Índice das Tabelas

Tabela 1. Distribuição dos imigrantes por sexo	41
Tabela 2. Distribuição dos imigrantes por nacionalidade	41
Tabela 3. Distribuição dos imigrantes por classes de Nacionalidade	42
Tabela 4. Distribuição dos imigrantes por nível escolaridade	44
Tabela 5. Distribuição dos imigrantes por idade	45
Tabela 6. Distribuição dos imigrantes por estado civil	47
Tabela 7. Distribuição dos imigrantes por motivo de entrada	48
Tabela 8. Distribuição dos imigrantes por profissão	49
Tabela 9. Distribuição dos imigrantes por tipo de celebração do casamento	50
Tabela 10. Distribuição dos imigrantes por proveniência da União Europeia	51
Tabela 11. Distribuição dos imigrantes por naturalidade	52
Tabela 12. Distribuição dos imigrantes por codificação da nacionalidade por classes	53
Tabela 13. Distribuição dos imigrantes por diferença entre naturalidade e nacionalidade	54
Tabela 14. Distribuição da naturalidade dos cônjuges	55
Tabela 15. Distribuição da nacionalidade dos cônjuges	56
Tabela 16. Distribuição por profissão dos cônjuges	57
Tabela 17. Estatística para idade dos cônjuges	58
Tabela 18. Frequência para números dos filhos do casamento	59
Tabela 19. Frequência para números dos filhos do outro casamento	60
Tabela 20. Os imigrantes que foram certificados na Câmara Municipal de Évora	65
Tabela 21. Relação entre ano de entrada em Évora e sua nacionalidade codificada (juntando União Europeia pelo SEF e Município de Évora)	65
Tabela 22. Sumários dos dados	84
Tabela 23. Valores p dos testes de Logrank e Peto para a comparação das curvas de sobrevivência para as diferentes variáveis em estudo	85
Tabela 24. Regressão de Cox para todas as variáveis dos dados dos imigrantes	97
Tabela 25. Modelo final do modelo de Cox	98
Tabela 46. Valores p do teste de Honell's	101
Tabela 27. Comparações dos valores dos coeficientes do modelo com e sem a observação 7	104
Tabela 28. Codificação dos Perfis	106

Tabela 29. Logaritmo da função $L(\theta)$ e AIC	109
Tabela 30. Teste Deviance	110
Tabela 31. Estimativas do modelo de fragilidade gama ajustados aos dados imigrantes	111

Resumo

Uma vez que existe uma grande necessidade de se obter um maior conhecimento das comunidades imigrantes residentes no Distrito, este trabalho tem como principal objectivo caracterizar a população de imigrantes comunitários residentes no distrito de Évora. Em particular, procuramos analisar os casamentos realizados não só entre os não nacionais e portugueses, bem como entre os não nacionais entre si.

Consideramos os 354 imigrantes que constituem os imigrantes comunitários inscritos legalmente nos Serviços Estrangeiros e Fronteiras de Évora entre 2006 a 2009. Por outro lado, analisamos, para o mesmo período, os casamentos dos imigrantes que casaram com os portugueses num total de 165 indivíduos.

Começamos por levar a efeito uma investigação básica de tipo descritivo para uma caracterização dos imigrantes relativamente a diferentes variáveis e investigamos algumas associações entre elas via tabelas de contingência. Refira-se que a maioria dos imigrantes é do sexo feminino, sendo a nacionalidade mais representativa a brasileira.

Com base no modelo de regressão de Cox identificamos factores de risco e diferentes perfis associados à rotura do casamento. Mostramos que há evidência estatística para considerar que um menor nível de escolaridade, o pertencer a um país da União Europeia e o ter entrado à procura de trabalho são níveis de factores que aumentam o risco de rotura do casamento.

Analisa-se, ainda, de uma forma crítica, a abordagem paramétrica, procurando modelar os dados através dos modelos Exponencial, Weibull, Lognormal e Log-logístico.

Abstract

There is a need to obtain a greater understanding of immigrant communities in the district; this work has as main objective to characterize the population of immigrant community residing in the district of Évora. In particular, we analyze not only marriages between Portuguese and non-nationals, as well as among non-nationals among themselves.

We consider the 354 immigrants who are legally registered immigrants in the Community Service of Foreigners and Frontiers of Évora, from 2006 to 2009. Moreover, we analyze, for the same period, the marriages of immigrants who intermarried with the Portuguese for a total of 165 individuals.

We start by carrying out a basic descriptive research in order to characterize the immigrants in relation to several variables and investigated some associations between them by contingency tables. It should be noted that most immigrants are women, and the more representative the Brazilian nationality.

Based on the Cox regression model, were possible to identify risk factors and identify profiles of high and low risk associated with the rupture of the marriage. We show that there is statistical evidence to conclude that a less educated levels, belonging to a country inside the European Union and have gone looking for work are factors that increase the risk of rupture of the marriage.

The parametric approach is analyzed also, in a critical way, seeking to model the data using exponential, Weibull, Lognormal and Log-logistic.

CAPÍTULO I

Introdução

1.1. Preliminares

Para a realização desta dissertação de Mestrado em Modelação Estatística e Análise de Dados, escolhi como tema “Uma caracterização dos imigrantes residentes no Alentejo”. Dentro de um grande leque de opções foi este que escolhi por ser um fenómeno complexo e actual. Ou seja, por ser um fenómeno sócio-demográfico e político, mas também humano e que está presente nos dias de hoje. Relativamente a este trabalho pretendo responder a certas dúvidas que possivelmente ainda subsistem acerca da imigração em Portugal especificamente no Alentejo. Para ir ao encontro dos objectivos a que me proponho realizei um estágio profissionalizante que procura recolher, analisar e discutir as informações que existem sobre o tema. Com este trabalho procuro caracterizar a população de imigrantes residentes neste distrito e, em particular, analisar os casamentos realizados não só entre os não nacionais e portugueses, bem como entre não nacionais entre si. Para este objectivo fui fazer estágio no SEF (Serviços de Estrangeiros e Fronteiras) em Évora e procurei rever a literatura sobre o assunto com a ajuda de livros de vários autores que escrevem sobre os imigrantes.

Refira-se que esta é uma problemática interessante e actual, uma vez que têm sido encontrados indícios de que os casamentos com indivíduos fora da nacionalidade de origem podem revelar estratégias para a obtenção da nacionalidade e, por outro lado, as relações parecem não ocorrer de forma idêntica para os homens e para as mulheres de uma mesma nacionalidade para alguns países de origem.

1.2. Contextualização do Problema

Os imigrantes são as populações que saem dos seus países de origem para outros países com objectivos de mudar de vida, procurar o trabalho, e alguns pretendem mudar as identidades das cidadanias. Também por causa de políticas de guerras que obrigavam as pessoas a sair para outros países que não tinham guerras, como Portugal. Actualmente, em Portugal há muitos imigrantes que se espalharam por todo o país, incluindo o distrito de Évora. O distrito de Évora é uma região que faz parte do Alentejo onde há muitos imigrantes de vários países como Brasil, Cabo Verde, Angola, Moçambique, Timor Leste, São Tomé e

Príncipe, Guiné Bissau, Índia, Espanha, Alemanha, entre outros. Neste trabalho procuramos caracterizar a população imigrante residente no Distrito de Évora, tomando como base os dados dos imigrantes que foram identificados pelo SEF (Serviços Estrangeiros e Fronteiras) do Distrito de Évora. Utilizaram-se os dados dos registos entre 2006 e 2009.

Os imigrantes comunitários são os imigrantes que fazem parte da União Europeia ou de origem da União Europeia. Por outro lado, esses imigrantes comunitários compostos também por imigrantes de países terceiros, por exemplo os imigrantes de fora da União Europeia, mas que têm relação com os Europeus.

Estudos sociológicos de caracterização dos imigrantes permitem identificar alguns aspectos positivos e negativos que se relacionem com as suas características. Os aspectos positivos são, por exemplo, as contribuições que os imigrantes dão para desenvolver o país e a sua cidade de origem, e integrarem-se bem como cidadãos na cidade de acolhimento. Os aspectos negativos, como os negócios feitos pelos imigrantes mas que não beneficiaram o país e as comunidades, os problemas criados com a sua vinda e integração, os distúrbios muitas vezes causados pelos próprios, entre outros.

No presente estudo começamos por levar a efeito uma investigação básica de tipo descritivo, utilizando uma metodologia que nos permite recolher as informações sobre perfis dos imigrantes do Distrito de Évora.

A Lei de base 23/2007, de 4 de Julho regula as condições e procedimentos de entrada, permanência, saída e afastamento de cidadãos estrangeiros do território português, bem como o estatuto de residente de longa duração¹. Com base nesta lei a entrada dos imigrantes em Portugal deve ser feita com os vários tipos de visto que existem, em que consistem as autorizações de residência, quais os trâmites de um processo de expulsão. Os imigrantes que precisam de tratar o processo para ter cartões de residentes e para poderem ter o direito de viver em Portugal tratam do processo no Serviço Estrangeiros e Fronteira de cada distrito.

Em relação ao tipo de medidas, a legislação portuguesa aplica as regras do Acordo de Schengen e das normas comunitárias, em relação ao segundo tipo, assegura aos imigrantes um conjunto significativo de direitos, ao nível do acesso à saúde, à justiça, à educação, à

¹Imigração em Portugal (página:7;2009)

segurança social e outros direitos fundamentais, decorrentes da Constituição e da Legislação que impedem discriminações por motivos de raça, cor, nacionalidade ou origem étnica.

Os cartões de residentes têm uma validade determinada, sendo que para definir a validade dos cartões é necessário saber quais são os tipos de imigrantes. Os cidadãos da União Europeia têm direito ao regime comunitário ao abrigo da Lei 37/2006 de 09/08, ao qual lhe é emitido um certificado de residência pela Autarquia Municipal válido por 5 anos renovável por cartão de residência por 10 anos. Aos nacionais de países terceiros é emitida uma autorização de residência ao abrigo da Lei 23/2007 de 04/07 válida por 1 ano e renovável por 2 anos.

Os comunitários são os imigrantes que fazem parte da União Europeia ou que têm uma relação formal com cidadãos da União Europeia. Por exemplo: casamento, trabalho, reagrupamento familiar, estudo, ser reformado. Os comunitários têm direito iguais aos portugueses, apenas não se podendo candidatar a Presidente da República.

Na realidade existem dois tipos de imigrantes que estão em Portugal: imigração legal e imigração ilegal. O imigrante legal tem os documentos legalizados e tem direito a viver ou ficar em Portugal durante um dado período de tempo, a imigração ilegal pode assentar em vários tipos como, por exemplo, associação, transportes ilegais, casamento por conveniência, agregação de mão-de-obra.

Imigrantes que vêm para Portugal algumas vezes procuram a legalização através do casamento com Portugueses sendo esses casamentos algumas vezes por conveniência só para legalizar os cartões de residentes. Com o estágio realizado nos Serviços de Estrangeiros e Fronteira do Distrito de Évora pretendeu-se caracterizar a população de imigrantes comunitários residentes no distrito. Uma vez que existe uma grande necessidade de se obter um maior conhecimento das comunidades imigrantes residentes nos diferentes distritos. Em particular, procuramos também analisar os casamentos realizados não só entre os não nacionais e portugueses, bem como entre os não nacionais entre si.

Inicialmente a ideia era estender o estudo ao Alentejo, mas questões de natureza logística, temporal e de disponibilização de dados levou a que o mesmo apenas ficasse restrito ao Distrito de Évora.

Houve, ainda, outras limitações ao nível das bases de dados, entre as quais destacamos:

- ❖ Os dados SEF existem em dossiers, não estando informatizados;
- ❖ Dificuldade em encontrar a data do fim do casamento;
- ❖ Muitas vezes os imigrantes que estavam casados com cidadãos nacionais separaram-se e foram renovar os cartões de residentes noutros distritos.

1.3. Actividades do Serviço de Estrangeiros e Fronteiras

Do que pudemos observar ao longo do período em que decorreu este estágio, podemos considerar que as actividades do SEF se podem dividir em dois grandes sectores: Documentário e Operacional.

1.3.1. Sector Documental

Este sector comporta fundamentalmente as seguintes actividades:

- a. Informações: as informações necessárias para dar aos imigrantes quando tem dificuldades de acesso ou tratar dos papéis. Para obter informações e impressos é possível fazê-lo através do sistema de agendamento on-line, do Portal do SEF ou pessoalmente no atendimento de cada delegação regional.
- b. Coimas Aplicadas: pela não renovação atempada do cartão, excesso, de permanência em território nacional, entre outras.
- c. Reagrupamento familiar: legalização de familiares (acendestes e descendentes) de cidadãos legais para viver em Portugal.
- d. Atendimento telefónico: para esclarecimento e informação aos utentes.
- e. Análise e instrução de processos: verificar e avaliar os documentos, e propor decisão de aceitar ou recusar o pedido.
- f. Atendimento de utentes: os estrangeiros que são atendidos ao balcão.
- g. Concessões de autorização residência/ primeiro cartão de residente emitido pelo SEF, que consiste na avaliação dos vistos concedidos pelas embaixadas, fazer o enquadramento legal da situação e mandar emitir o cartão em Lisboa.
- h. Renovação de autorização residência: cada cartão tem uma terminada validade, obriga os estrangeiros a fazer a sua renovação conforme cada situação: estudo, trabalho, reagrupamento familiar, voluntariado, entre outros.
- i. Entrada e saída de correspondência relacionada com a actividade do SEF.
- j. Elaboração contabilidade referente ao pagamento dos cartões de residentes.

- k. Estatística: consequência do investimento do SEF na melhoria da qualidade da informação estatística, em cada ano é possível recolher dados extraídos exclusivamente do Sistema Integrado de Informação do SEF sobre a população estrangeira residente legal em Portugal.
- l. Arquivo dos processos que foram emitidos.

1.3.2. Sector Operacional

Este sector é exercido no terreno e comporta fundamentalmente as seguintes actividades:

a. Fiscalização de imigrantes

No domínio do controlo e fiscalização da permanência e actividades de estrangeiros em território nacional, o SEF concretiza as atribuições que lhe estão legalmente confiadas, centrando a sua actuação na repressão das redes de recrutamento de mão-de-obra ilegal e de tráfico de seres humanos. Como pressuposto desta actuação, o reforço da fiscalização e da investigação criminal constituem um factor de equilíbrio tendo em vista uma imigração mais regulada e tutelada dos direitos fundamentais dos imigrantes. Actuação conjunta com outras entidades, nomeadamente, GNR, PSP, PJ, ACT, Segurança Social e o Corpo Nacional de Polícia de Espanha, este no âmbito da cooperação, luso-espanhola. Ver se o imigrante é legal ou ilegal; se for ilegal poderá ser detido e voltar para o país origem.

b. Investigação criminal

O SEF possui competência para proceder à averiguação e investigação criminal de crimes de auxílio à imigração ilegal e outros com ele conexos, nomeadamente, do crime de tráfico de pessoas, sem prejuízo das atribuições de outras entidades neste domínio. Não deverá descuidar-se que o combate às redes de imigração e de mão-de-obra ilegais constituem um objectivo estratégico da actuação do Serviço. O casamento de conveniência e angariação de mão-de-obra ilegal são exemplos.

c. Detecção de imigrantes

Na concretização das suas atribuições no âmbito do controlo de fronteiras e fiscalização de estrangeiros, o SEF desenvolve actividades de identificação e

peritagem documental, em especial no que se prende com documentos de viagem e de identidade.

O reforço da segurança documental e a introdução de dados biométricos em documentos de viagem Europeus concorrem decisivamente para a redução da falsificação deste tipo de documentos. Por outro lado, o alargamento da União e da sua fronteira externa, constituiu um factor de deslocação da utilização de documentos fraudulentos, tendência já assinalada noutros Estados Membros.

O Serviço de Estrangeiros e Fronteiras possui competência para vigiar e fiscalizar a circulação de pessoas nos postos de fronteira, incluindo a zona internacional dos portos e aeroportos. Durante este processo verifica-se as nacionalidades dos documentos, nacionalidades dos portadores e proveniência ou destino dos documentos.

d. Outros

Podem ainda referir-se, entre outros, processos de expulsão com notificação de abandono de território nacional; operações móveis de fiscalização de ruas, escoltar cidadãos estrangeiros no território nacional (levá-los para centros de instalação temporária, tribunal) escoltar cidadãos estrangeiros ao país de origem; e Pco's (Processo de coimas ou multas aplicadas).

1.4. Actividades do Estágio no SEF

Estágio é um trabalho em que se aprende a aplicar conhecimentos adquiridos na fase curricular, permitindo dar uma experiência profissional ao estagiário. A actividade desenvolvida no SEF ao longo deste estágio foi integrada no sector documental.

A minha integração no grupo de trabalho foi boa não desistindo de aprofundar o meu conhecimento. Pude constatar a forma como os empregados do SEF tratava da documentação necessária para resolver as diversas questões que apareciam diariamente.

A actividade desenvolvida durante o período do estágio foi a seguinte:

- a. Notificação dos requerentes para levantamento dos cartões da residência imigrantes;

- b. Atendimento telefónico: chamadas telefónicas para informar os requerentes da data que podia vir levantar os cartões residentes ou título da residência;
- c. Observação do atendimento geral;
- d. Recolha os dados dos imigrantes comunitários nos anos de 2006 a 2009 para o trabalho de pesquisa.

Finalmente este estágio orientou-se para uma componente de investigação, nomeadamente a realização do trabalho de mestrado, de forma a aprofundar os conhecimentos adquiridos. Para além disso o estágio permitiu aprender o tratamento dos processos documentais dos imigrantes.

No capítulo I apresenta-se a introdução que é composta por caracterização do problema, objectivos, limitações do estudo e as actividades do SEF. De seguida no capítulo II apresenta-se tópicos gerais e metodologia. A caracterização dos imigrantes do distrito de Évora é apresentada no capítulo III. No capítulo IV apresenta-se o estudo de estatístico do casamento dos imigrantes. Por último as conclusões e recomendações são apresentadas no capítulo V.

CAPÍTULO II

Tópicos Gerais e Metodologia

2.1. A Imigração

2.1.1. Conceito

A imigração é uma maneira de migração, de intenção das pessoas que querem mudar para outro país novo, em função dos seus objectivos. Portanto, a maioria das pessoas que se deslocam de um país para outro não pretende deixar a própria terra permanentemente. De acordo com Malheiros (1996), a imigração é o processo de entrada num país ou região com o objectivo de fixação, temporária ou definitiva, nessa área². Por isso considera-se que a imigração é o resultado de um processo das decisões individuais ou familiares, mas também faz parte de um processo social. Em termos económicos, a migração é tanto um fenómeno mundial como o comércio de mercadorias. Designa o movimento das populações, mas faz parte de um modelo mais profundo e é um sinal de relações económicas, sociais e culturais em transformação.

Segundo a equipa do SOS Racismo (2005), os imigrantes são pessoas que entram num determinado território, com o objectivo de mudar a sua residência para esse local³. A palavra imigração refere-se a um movimento entre os indivíduos do seu país de origem para outro país, a imigração refere-se ao movimento de entrada em dado país, região, cidade. Com estes conceitos podemos dizer que os imigrantes são as pessoas que se movimentam para entrar num país novo ou noutra país e trouxeram os seus objectivos assinalados, por exemplo: os refugiados foram expulsos pela guerra civil, tragédias naturais, perseguição, procurar segurança fora dos seus países, saem para o trabalho temporário e estudar. Considera-se como imigração o movimento de entrada, com âmbito permanente ou temporário e com a intenção de trabalho e/ou residência, de pessoas ou populações, de um país para outro.

Os imigrantes comunitários são os imigrantes que fazem parte da União Europeia, por outro lado existem os imigrantes com a sua nacionalidade de países terceiros, mas quando têm relação com os Europeus esses imigrantes também fazem parte dos

² Malheiros, J, Macaísta. Imigrantes na Região de Lisboa (Os anos da Mudança), 1996, p:30

³ Racismo, SOS. Imigração e Etnicidade (Vivências e Trajectórias de Mulheres em Portugal), 2005, p: 29

comunitários. Ao cidadão comunitário é admitida a sua entrada em território nacional mediante a apresentação de bilhete de identidade ou passaporte válidos⁴.

2.1.2. Razões da Imigração

Os imigrantes para entrarem num país que não é o seu país de origem para viver ou estudar, trabalhar, entre outras, precisam de expor as suas razões. As razões que levam as pessoas a saírem do seu país de origem, compreendendo e referindo algumas políticas que sustentam este fenómeno. Não se deve confundir a figura do imigrante com a do turista, que ingressa num país apenas com o intuito de visitá-lo e depois retornar ao seu país natal⁵.

As razões da vinda para Portugal podem ser por motivos de férias, de estudo, de trabalho e melhorar as condições de vida⁶. Relativamente aos motivos que trouxeram os imigrantes são geralmente trabalho, reagrupamento familiar, estudo, reforma, guerra e outros. Estas são as razões que motivaram a maioria dos imigrantes a deixarem o seu país de origem e ir para outro país. Assim os motivos que justificam a atracção exercida por imigrantes comunitários encontram-se clarificados e a motivação das pessoas para abandonarem o seu país de origem e imigrar para a Portugal.

As condições de vida são um motivo para tal, a miséria em que vivem os imigrantes no seu país, o desemprego. Todos estes factores são verdadeiras razões para as pessoas abandonarem o seu país; mas é evidente que não basta dizer que estes fenómenos são os mais importantes, pois é necessário reflectir sobre os conflitos político-militares que também provocam tais fluxos migratórios.

De acordo com o relatório anual do SEF, outra das razões apontadas para a imigração para este país foi a existência de familiares em Portugal os quais serviram de sustento no período inicial, oferecendo alojamento, trabalho e criando oportunidades para a estabilidade económica do imigrante⁷.

Ao longo deste estudo social que é a imigração, revelaram-se as razões que levam as pessoas a saírem do seu país de origem e a imigrarem para um país distante e desconhecido. Entretanto, os imigrantes ao longo de todo este processo, passam por situações que nós não

⁴ Meireles. J, Meireles. M, Barbosa. M. GUIA PRÁTICO JURÍDICO DO CIDADÃO IMIGRANTE (NÃO COMUNITÁRIO): DIREITOS E DEVERES EM PORTUGAL, 2002, p. 41

⁵ <http://pt.wikipedia.org/wiki/Imigração>

⁶ Ibid. p; 208

⁷ SEF, relatório anual de 2007.

nos apercebemos e nem sequer imaginamos; isto leva a deduzir as consequências de tal fenómeno⁸.

Uma das consequências da imigração é o aproveitamento dos proprietários de empresas, fábricas, e de obras. Os imigrantes são aproveitados como “homens de mão-de-obra clandestina”, que agrupam uma grande quantidade de mão-de-obra, ilegal ou não, e que por fim não lhes pagam as remunerações devidas. Portanto, as razões que explicam as migrações são inúmeras (político – ideológicas, étnico – raciais, profissionais, económicas, catástrofes naturais etc.), embora as razões económicas sejam as predominantes.

2.1.3. Política de Imigração

A política de imigração é um aspecto definido legalmente que pode regular a sobrevivência dos imigrantes num país, e esta política é muito importante para tratar a entrada e saída dos imigrantes. Por exemplo, em Portugal o Governo com o Decreto-Lei N.º 244/98, de 08 de Agosto, instituiu um novo regime jurídico da política da imigração. E este Decreto aplica-se regularmente condições de entrada, permanência e afastamento dos imigrantes do território português (Falcão, 2002)⁹.

Portanto, o fenómeno crescente da imigração, cada vez mais visível na sociedade portuguesa, obrigou o Estado a desenvolver, a partir de meados da década de noventa, políticas de intervenção ao nível da imigração e políticas de direitos e deveres dos imigrantes. Países, como Portugal, até ainda há pouco tempo países «exportadores» de mão-de-obra estão pouco habituados à recepção de imigrantes. Não é pois de estranhar que este tipo de Estados não tenha desenvolvido qualquer plano de integração destes cidadãos, até muito recentemente, abandonando-os a todos os tipos de exploração.

Podem-se considerar como «políticas de imigração» dois tipos de medidas: legislativas e regulamentares. Primeiramente, as medidas relativas à entrada de estrangeiros no país e, em seguida, as medidas destinadas a promoverem a inserção ou a integração de grupos ou comunidades estrangeiras residentes em Portugal.

Em relação às medidas de integração na sociedade portuguesa, a situação é no momento actual globalmente positiva. Em termos constitucionais, tem-se lutado para

⁸ <http://sosracismo.blogspot.com/>

⁹ Falcão. Luísa (2002), A imigração em Portugal – Relatório síntese elaborado pelo Delta Consultores, Lisboa, p10.

estabelecer o princípio da igualdade, o da não discriminação dos cidadãos e o princípio da equiparação de direitos entre nacionais e estrangeiros. Esses direitos incluem o do acesso à educação, à saúde, à protecção social, aos tribunais, além de todo o conjunto de direitos fundamentais relativos à vida, à família e à segurança pessoal (Trindade, 2001). Os imigrantes legalmente residentes em Portugal gozam da generalidade dos direitos, liberdades e garantias pessoais (tais como o direito à vida, integridade e identidade pessoal, liberdade e segurança), de participação política e dos mesmos direitos que os trabalhadores nacionais, assim como dos direitos e deveres económicos, sociais e culturais. Os trabalhadores imigrantes têm direito, sem distinção de nacionalidade, «raça» ou território de origem, à segurança social, à retribuição do trabalho, segundo a quantidade, natureza e qualidade; à organização do trabalho em condições socialmente dignificantes, à prestação do trabalho em condições de higiene, segurança e saúde; ao repouso e aos lazeres e a um limite máximo da jornada de trabalho, entre outros. A lei que institucionalizou o Rendimento Mínimo Garantido previu que esse direito seja reconhecido aos residentes legais, incluindo os imigrantes, o que significa o reconhecimento da sua cidadania.

Por esta altura, o único organismo encarregue da temática da imigração era o Serviço de Estrangeiros e Fronteiras (SEF), na dependência do Ministério da Administração Interna. No entanto, as suas competências (direccionadas para o controlo de fronteiras, para a documentação e regularização de estrangeiros, fiscalização das suas condições de permanência, prevenção e combate à criminalidade associada ao fenómeno migratório), embora importantes para a integração dos imigrantes, não conseguiam (nem podiam) dar resposta a todas as exigências delas decorrentes¹⁰.

Os cidadãos provenientes de países fora da União Europeia são os imigrantes que fazem parte de países terceiros. Quando esses imigrantes entram em Portugal deve-se:

- ❖ Verificar se tem um documento de viagem válido reconhecido (**passaporte**) (n.º 1, art.º 12.º e art.º 59.º da LI);
- ❖ Verificar se tem um **visto de entrada** válido e adequado à finalidade da deslocação (n.º 1, art.º 13.º da LI);
- ❖ Verificar se tem **meios de subsistência suficientes** para o período que pretende permanecer em Portugal (n.º 1, art.º 14.º da LI) Pode ser contudo dispensado desta

¹⁰ [http://www.sef.pt/documentos/56/VersaoFinal, O ImpactodaImigracaonasSociedadesdaEuropav1Port.pdf](http://www.sef.pt/documentos/56/VersaoFinal,O%20ImpactodaImigracaonasSociedadesdaEuropav1Port.pdf)

exigência se provar ter alimentação e alojamentos assegurados durante a respectiva estadia (Portaria n.º 1426/98, de 11/12, do Ministro da Administração Interna);

- ❖ Verificar se tem documentos que justifiquem o motivo e as condições da sua permanência em Portugal (art.º 15.º da LI)¹¹.

Portanto os imigrantes que fazem parte da União Europeia e os que não fazem parte da União Europeia têm diferentes critérios para tratar os papéis para a legalização em Portugal. Aos imigrantes comunitários basta apresentar bilhete de identidade ou passaporte válido para entrar em Portugal.

Segundo a equipa do SOS Racismo (2002), pode haver várias razões para a deportação dos imigrantes, tais como, posse de drogas com intenção de distribuir, alvo de várias condenações, assaltos, permanência ilegal, homicídio e violência doméstica¹².

Entretanto, a política de imigração é um ambiente ou espaço em que pode permitir aos imigrantes poderem imigrar ou viver noutra país. Esta política é definida pelas autoridades do País e dentro desta política há alguns critérios ou leis que tratam os direitos dos imigrantes. Quando os imigrantes não cumprem estas regras deportam-se para os seus países de origem. Por outro lado, os que cumprem estas políticas podem viver à vontade como cidadãos legais.

2.1.4. Casamentos dos Imigrantes

Como em qualquer processo histórico desta natureza há um factor que é essencial, mas que é sempre doloroso, o tempo. É necessário tempo para que ocorram cruzamentos de povos e culturas, para que entre pessoas de origens muito diferentes se crie e consolide uma história em comum, permitindo o nascimento de novas culturas que superem velhos ódios e discriminações¹³. Relacionado com esta teoria há vários cruzamentos que ocorrem entre os imigrantes, entre vários aspectos especificamente são os casamentos dos imigrantes com os cidadãos nacionais.

¹¹ Meireles. J, Meireles. M, Barbosa. M. GUIA PRÁTICO JURÍDICO DO CIDADÃO IMIGRANTE (NÃO COMUNITÁRIO): DIREITOS E DEVERES EM PORTUGAL, 2002, p. 42

¹² Racismo SOS, 2002

¹³ <http://confrontos.no.sapo.pt/page9b.html>

De acordo com Michel (1983), o casamento é somente um dos múltiplos aspectos de troca entre grupos humanos que podem contrair aliança, graças às regras de exogamia que interdita o casamento com um membro da família legal¹⁴. O casamento pode definir-se como um contrato celebrado entre duas pessoas de sexo diferente, que vivem ou passam a viver em comum e que na maioria das vezes tem como símbolo a troca de alianças. O casamento entre imigrantes e cidadãos nacionais assume diferentes tendências, consoante os noivos sejam africanos, brasileiros e ucranianos. Os brasileiros (as) saem mais do seu grupo para casar com os portugueses (as).

Há várias razões que levam os imigrantes a casar com os portugueses (as), um deles é o casamento por conveniência. Este tipo de casamento visa beneficiar os imigrantes para obterem o título de residência para viver em Portugal. Por exemplo, a lei permite que quando um estrangeiro (imigrante) case com um português obtenha automaticamente a autorização de residência e pode pedir a nacionalidade ao fim de três anos. Mas, pelo contrário, se um português que case com um imigrante ilegal para que este obtenha a autorização de residência ou a nacionalidade portuguesa pode ser condenado de um a quatro anos de prisão. O SEF da região Norte detectou 25 a 30 imigrantes de casamentos de conveniência por mês¹⁵.

2.2. Metodologia do Trabalho

2.2.1. Local e Tempo de Estágio

Como referi anteriormente, este estágio decorreu no Serviço Estrangeiros e Fronteiras Évora, entre o dia 1 de Março de 2010 e o dia 31 de Julho de 2010.

2.2.2. População e Amostra

A população é basicamente um conjunto de indivíduos ou objectos em uma região ou de um determinado local que se torna o alvo da investigação. Com base na definição acima pode concluir a população é constituída por todos os imigrantes que fixaram residência no distrito de Évora nos anos de 2006 a 2009 que são 2391 indivíduos.

A base de amostragem foi aplicar um corte só para os imigrantes comunitários que entraram, conforme a base de dados considerada, sendo constituída pelos imigrantes

¹⁴ Michel, Andrée, Sociologia da Família e do Casamento, Coimbra, 1983, p. 58

¹⁵ Notícia Expresso: <http://aeiou.expresso.pt/imigracao-sefnorte-detetou-25-a-30-casamentos-de-conveniencia-por-mes=f596521>

inscritos nos Serviços Estrangeiros Fronteiras nos anos de 2006 até 2009. As informações necessárias foram obtidas através do inquérito dos processos comunitários no SEF. Com base nos dados que recolhi entre 2006 e 2009 entraram 354 imigrantes comunitários em Évora. Neste trabalho analisei os casamentos em 165 indivíduos.

2.2.3. Variáveis de Investigação

Neste trabalho foram consideradas diversas variáveis medidas em diferentes escalas: nacionalidade dos imigrantes e do seu cônjuge, idade, sexo, estado civil, designação da morada, os motivos de entrada, profissões dos imigrantes e do seu cônjuge, nível escolaridade, residência antes e depois de casamento, naturalidade dos imigrantes e dos seus cônjuges, tipo de celebração do casamento, números de filhos do casamento e filhos de outros casamentos, Distrito, União Europeia, tempo de duração do casamento.

2.2.4. Técnica de Recolha de Dados

A técnica de recolha de dados deste estudo foi a análise documental. Utilizei a documentação para obter os dados dos imigrantes em Évora no período de 2006 a 2009. A documentação é "dedicado a obter dados directamente a partir da pesquisa, incluindo livros, regulamentos, relatórios sobre as actividades, fotografias, documentários, dados de pesquisa relevantes".

O Trabalho foi realizado com base em dados recolhidos a partir dos inquéritos por questionário que foram aplicados pelo SEF aos imigrantes que foram tratar os cartões Residentes nos Serviços Estrangeiros e Fronteiras nos anos de 2006 a 2009 no distrito de Évora. Os dados não se encontram tratados e neste trabalho apenas vão ser tratados os dados relativos aos processos classificados como comunitários.

2.2.5. Técnica de Análise dos Dados

Em cada estudo, após os dados colectados, os próximos passos a serem feitos por um investigador é a análise e discussão dos dados. Com base no tamanho das amostras colhidas pelos investigadores neste estudo, os processos de análise dos dados são os que a seguir se descrevem.

2.3. Análise Exploratória de Dados

Em qualquer estudo a análise exploratória de dados é muito importante, na medida em que torna possível produzir um estudo descritivo e interpretar resultados, resumir a informação contida num conjunto de dados, através do recurso a tabelas, gráficos e diversas medidas descritivas, de forma a evidenciar os aspectos mais relevantes ou pertinentes, produzir uma descrição simples, concisa e completa de uma população caracterizada por um atributo. De acordo com Maroco (2007), a análise exploratória de dados é uma análise composta por duas partes, uma ou poucas variáveis e múltiplas variáveis¹⁶. Na análise com uma ou poucas variáveis estuda-se a análise descritiva que é necessária para caracterizar a amostra e para determinar as chamadas medidas de tendência central como média, mediana e outras (que procuram caracterizar o valor de variável sob estudo que ocorre com mais frequências) e a dispersão das observações em torno das estatísticas de tendência central ou na amostra. Ainda na análise exploratória podemos usar as medidas da forma da distribuição dos elementos das amostras – as ditas medidas de assimetria e achatamento. Finalmente, no caso das amostras multivariadas podemos utilizar as medidas de associação que caracterizam a intensidade e a direcção da variação comum entre variáveis.

Fundamentalmente, construímos e apresentámos as tabelas de frequência referentes às questões apresentadas no questionário e, quando foi caso disso, dos dados recodificados. Aplica-se análise estatística descritiva que pode caracterizar os perfis dos imigrantes comunitários no distrito de Évora. Para o efeito, utilizámos o programa Excel e os software de análise estatística SPSS e R.

Na parte de investigação o primeiro objectivo é descrever o fenómeno, por isso a primeira fase da análise dos dados utiliza-se a análise univariada, elaborando tabelas de frequências e o cálculo das medidas de localização, dispersão, assimetria e achatamento para as variáveis em estudo. Portanto, na análise univariada cada variável é estudada isoladamente.

No entanto, as medidas que estão explicadas em cima não poderão ser calculadas para qualquer tipo de variável, tudo isso depende da escala de valores de cada variável. Assim tem um passo mais importante é clarificarmos as variáveis quanto à sua escala, portanto esta classificação é de importância crucial para utilizar de forma adequada a estatística.

¹⁶ Maroco, J. (2007), *Análise Estatística – Com Utilização do SPSS*, Edições Sílabo, Lda, Lisboa, p.37.

Além de aplicar análise exploratória, averigua-se também a associação entre algumas variáveis porque várias variáveis têm escala nominal, as medidas de associação também designadas por coeficientes de correlação que são frequentemente usadas, incorrectamente diga-se, para inferir sobre relações causais (Maroco, 2007)¹⁷. De seguida, procuramos verificar até que ponto os resultados implicam associação forte entre variáveis, calculadas através de simples subtracções e risco relativo e odds ratio.

2.4. Análise de Sobrevivência

A análise de sobrevivência é umas das áreas da estatística mais importantes em investigação, na análise de eventos que ocorrem uma única vez. Os eventos podem ocorrer uma única vez ou diversas vezes e os eventos são atributos especiais, porque estão sempre associados ao tempo. De acordo com Oliveira, (2009) um evento é uma variável binária que está associada ao tempo que decorreu desde o início da observação do indivíduo até ao momento em que ocorreu o evento¹⁸. No nosso estudo observa-se a duração do casamento até à ocorrência do divórcio (falha de casamento).

Colosimo e Giolo (2006) referem que na análise de sobrevivência a variável resposta é, geralmente, o tempo até ocorrência de um evento de interesse. Este tempo é denominado tempo de falha, podendo ser o tempo até a morte do paciente, bem como até a cura ou recidiva de uma doença¹⁹. Por causa da sua dependência do tempo, precisamos de métodos estatísticos especiais para analisar este tipo de variáveis. Um das características dos dados de sobrevivência são a existência de censura.

A análise sobrevivência é uma colecção de procedimento estatístico para análise de dados cuja variável resposta é o tempo até ocorrência de um evento de interesse. Este tempo denominado tempo de falha, de sobrevivência ou de vida. Esse tempo de falha pode ser o tempo até a separação de um indivíduo casado.

Os dados de sobrevivência são caracterizados pela presença de censura. Nesse contexto, censura corresponde às observações incompletas ou parciais sobre os indivíduos. A ocorrência de dados censurados informa que o tempo de falha foi superior ao tempo observado no estudo. Todos os dados, até mesmo os censurados, devem ser considerados na análise de sobrevivência, pois mesmo incompletas essas observações possuem informações

¹⁷ Ibid: 42.

¹⁸ Oliveira. A. G, Bioestatística Epidemiologia e Investigação, 2009, p. 138.

¹⁹ Colosimo.E. A e Giolo. S.R, 2006, Análise Sobrevivência Aplicada, 1ª edição, Edgar Blucher, Brasil

sobre o tempo de vida do indivíduo e a não considerações das censuras no cálculo estatístico podem conduzir a conclusões viciadas (Colosimo & Giolo, 2006). Portanto neste estudo aplica-se a análise de sobrevivência procuramos modelar o tempo de “sobrevivência do casamento” em função de diversas covariáveis de modo a identificar factores de risco e factores protectores.

Os objectivos de uma análise estatística envolvendo dados de sobrevivência estão geralmente relacionados, em medicina, com a identificação de factores prognóstico para uma certa doença ou com a comparação de tratamentos em estudo clínico enquanto controlado por outros factores²⁰. Neste estudo, com a análise de sobrevivência procuramos modelar o tempo de “sobrevivência do casamento” em função de diversas covariáveis de modo a identificar factores de risco e factores protectores.

Na análise de sobrevivência é necessária a existência de indivíduos observados durante um determinado período de tempo, muito frequentemente não é possível obter em todos os indivíduos dados do tempo até ao evento. Como já tínhamos referido os eventos codificam-se como uma binária, que indica se ou evento ocorreu ou não, em que conta o tempo até ao evento. Se o estudo ainda não acabou mas o indivíduo abandonou o estudo e o evento não chegou a ocorrer ou o estudo terminou antes de evento ocorrer são chamados dados censurados.

As principais grandezas estudadas são a função de sobrevivência e a função hazard. A primeira descreve a distribuição dos tempos até a ocorrência do evento e a segunda refere-se a um potencial instantâneo da ocorrência do evento dado “sobrevivência” até um tempo arbitrário t isto é, dado que não se realizou o evento até ao tempo t .

Do ponto de vista matemático, essas quantidades podem ser definidas da seguinte maneira. Seja T o tempo até a ocorrência do evento. Então, para um dado tempo t a função de sobrevivência $S(t)$ é a probabilidade de que o tempo até a ocorrência do evento seja maior ou igual a t , ou seja,

$$S(t) = P(T > t)$$

Desta maneira, $S(t)$ descreve a probabilidade de que evento dure mais tempo do que t . A função hazard, $h(t)$, tem uma definição matemática um pouco menos intuitiva.

²⁰ Ibid, Pg: 29

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

$h(t)$ descreve a taxa condicional de ocorrência de evento num determinado instante de tempo t dado sobrevivência até imediatamente antes do tempo t .

Para a estimação da função sobrevivência têm-se os estimadores não paramétricos de Kaplan-Meier, tabela de vida ou actuarial e o estimador de Nelson Aalen (Colosimo & Giolo, 2006).

Os estimadores diferenciam-se no número de intervalos utilizados para construção de cada um deles, sendo que o estimador de Kaplan-Meier e o de Nelson-Aalen baseiam-se em números de intervalos iguais ao número de falhas distintas. Para o estimador de tabela de vida ou actuarial, os tempos são agrupados de forma arbitrária. O estimador de Kaplan-Meier é o mais utilizado para estimar $S(t)$ em análise sobrevivência. O método de Nelson-Aalen é mais recente do que o Kaplan-Meier e foi utilizado em informações provenientes de censos demográficos para estimar características associadas ao tempo de vida dos seres humanos. O estimador da tabela de vida ou actuarial consiste na construção de uma tabela de vida dividindo o eixo do tempo em um certo número de intervalos. A Tabela de Vida ou Método Actuarial é uma das mais antigas técnicas estatísticas utilizadas para estimar características associadas à distribuição dos tempos de falha.

De acordo com Oliveira (2009), o método actuarial estima a probabilidade cumulativa de não ocorrer o evento em cada período de tempo. Se não houvesse dados censurados, isto é, se todos os indivíduos de estudo tivessem sido observados até ocorrer o evento, a probabilidade cumulativa de não ocorrer o evento seria obtida dividindo, em cada intervalo, o total de indivíduos em que não ocorrera ainda o evento pelo total da amostra.

2.4.1. O Estimador de Kaplan-Meier

Para Oliveira (2009), o método de Kaplan-Meier é idêntico ao método actuarial, com a única diferença de os intervalos de tempo serem definidos pela unidade de tempo utilizada para contar o tempo de observação. As estimativas das probabilidades cumulativas constituem a função de sobrevivência, habitualmente apresentada pela **curva de Kaplan-Meier**²¹. O estimador de Kaplan-Meier é um estimador não paramétrico da função de sobrevivência, que considera tantos intervalos quantos forem o número de falhas distintas, o

²¹ Oliveira. A. G, Bioestatística Epidemiologia e Investigação, 2009, pg: 142

que possibilita maior precisão nos cálculos. Há estudos que mostram a superioridade do estimador de Kaplan-Meier, de modo que esse tem sido o mais utilizado em pesquisas clínicas e vem ganhando mais espaço na área fiabilidade (Colosimo & Giolo, 2006).

O estimador de Kaplan-Meier é uma importante ferramenta para a análise de dados censurados, visto que é facilmente calculado e requer poucas suposições (Machin, Cheung, Parmar, 2006). Este também é conhecido como estimador produto-limite, e este estimador é uma adaptação da função sobrevivência empírica que, na ausência de censuras, é definida como:

$$\hat{S}(t) = \frac{\text{no. de observações que não falharam até o tempo } t}{\text{no. total de observações no estudo}}$$

$\hat{S}(t)$ é uma função escada com degraus nos tempos observados de falha de tamanho $1/n$, em que n é o tamanho da amostra. Se existirem empates em um certo de tempo t , o tamanho do degrau fica multiplicado pelo numérico de empates²².

Para definição do estimador de Kaplan-Meier, consideramos, conforme descrito por Colosimo e Giolo (2006), que existem n pacientes no estudo e k ($\leq n$) falhas nos tempos distintos e ordenados de falha, isto é $t_1 < t_2 < \dots < t_j$. Considera-se $S(t)$ uma função discreta com probabilidade maior do que zero somente nos tempos de falha t_j , $j = 1, 2, \dots, k$ tem-se que:

$$S(t_j) = (1 - q_1)(1 - q_2) \dots (1 - q_j)$$

Onde q_j é a probabilidade do indivíduo morrer no intervalo $[t_{j-1}, t_j)$, sabendo que ele não morreu até t_{j-1} e considerando $t_0 = 0$, isto é

$$q_j = P(T \in [t_{j-1}, t_j) | T \geq t_{j-1})$$

Assim estimado por:

$$q_j = \frac{d_j}{n_j}$$

²² Colosimo.E. A e Giolo. S.R, 2006, Análise Sobrevivência Aplicada, 1ª edição, Edgar Blucher, Brasil, p.35.

Onde d_j é o número de falhas (ou mortes) em t_j para $j = 1, 2, 3, \dots, k$ e n_j é o número de indivíduos sob risco em t_j , isto é, aqueles que não falharem não foram censurados até o instante anterior até t_j .

O Estimador de Kaplan-Meier na sua construção considera tantos intervalos de tempo, quantos forem o número de falhas distintas. Ele é definido da seguinte forma:

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

em que $t_1 < t_2, \dots, t_k$, e k representam os tempos distintos em ocorrência do evento

$\hat{S}(t)$ é uma função com degraus nos tempos observados.

d_j é o número de eventos em t_j , $j = 1, \dots, k$

n_j é o número de indivíduos sob risco em t_j , ou seja, os indivíduos em que não ocorreu o evento e não foram censurados até ao instante imediatamente anterior a t_j ²³.

Neste caso estamos a observar as falhas de um determinado casamento, ou seja, no instante que ocorreu um evento (divórcio ou separação).

A partir da curva de Kaplan-Meier é possível obter-se estimativas de percentis. Pode-se também obter o valor de interesse directamente das estimativas de Kaplan-Meier, ou por interpolação, caso o valor do tempo de interesse, por exemplo, o tempo mediano, estiver ao longo de grau da curva de Kaplan-Meier (Colosimo & Giolo, 2006). Segundo Oliveira (2009), o tempo mediano é definido como o menor tempo observado, para o qual o valor da função de sobrevivência estimada é menor ou igual a 0,5. Mas, como a curva de Kaplan-Meier é uma função escada, Colosimo e Giolo (2006) destacam que a estimativa mais adequada para a mediana é obtida por interpolação.

Além do estudo descritivo dos dados e de construir as curvas de sobrevivência, pode-se ter como o objectivo a comparação de vários tratamentos. Isto é, possivelmente existirá para cada tratamento uma curva de sobrevivência que, posteriormente, serão comparadas entre si. Para comparação de duas ou mais curvas existem alguns testes não paramétricos, são teste **Log-rank**, teste de **Breslow** e teste de **Wilcoxon**.

²³ Ibid.: 37

Para comparação dos tratamentos ou das covariáveis foi utilizado o teste Logrank. Inicialmente verificou-se a hipótese de igualdade de todos os tratamentos. Sendo constatada a presença de diferenças entre os tratamentos, identificaram quais tratamentos diferiam entre si. Para identificar os tratamentos que diferiam entre si foram realizadas comparações entre tratamentos. O teste Logrank utiliza-se para comparação de duas ou mais curvas de sobrevivência, é dado como segue: suponha que g ($g \geq 2$) funções de sobrevivência devam ser comparadas. Seja U a estatística que compara o número de falhas (separadas) nos $1, 2, \dots, g-1$ grupos com seus respectivos valores esperados. Matematicamente, temos:

$$U_{LK} = \sum_{j=1}^r (d_{kj} - n_{kj} d_j n_j^{-1})$$

em que d_{kj} é o número total de falha, n_{kj} é o número de sobreviventes e r é números dos indivíduos sob risco. Para $k = 1, 2, \dots, g-1$, A quantidade U_{LK} será expressa na forma de um vector com $(g-1)$ componentes, que será denotada por U_L . A covariância entre U_L e U_{LK} é dada por :

$$V_{LKK'} = \sum_{j=1}^r \frac{n_{kj} d_j (n_j - d_j)}{n_j (n_j - 1)} \left(\delta_{kk'} - \frac{n_{kj}}{n_j} \right)$$

Para $k, k' = 1, 2, \dots, g-1$ onde $\delta_{kk'}$ é tal que :

$$\delta_{kk'} = 1 \text{ se } k = k' \text{ e}$$

0 se $k \neq k'$

Que sob a hipótese nula tem uma distribuição com $(g-1)$ graus de liberdade. O teste **Log-rank** permite-nos testar a hipótese nula de um evento ocorrer com igual frequência em populações diferentes²⁴. O teste de Log-rank faz comparações com várias curvas e utiliza o teste do qui-quadrado, que consiste nas comparações do número de eventos observados em cada grupo com o número esperado se a função de risco for igual para todos. De acordo com Colosimo e Giolo (2006), o teste Log-rank é muito utilizado em análise de sobrevivência e é particularmente apropriado quando a razão das funções de risco dos grupos a serem comparados é aproximadamente constante²⁵. O teste Log-rank é o método mais popular de comparar a sobrevivência dos grupos e podemos utilizar o teste Log-rank quando as curvas

²⁴ Ibid, pg: 144

²⁵ Ibid, pg: 55

de sobrevivência não se cruzam e quando as curvas se cruzam utilizamos o teste Breslow ou o de Wilcoxon. Existem também outros testes estatísticos não paramétricos que foram propostos para comparar funções de sobrevivência. No caso particular da comparação de duas funções de sobrevivência, dada por:

$$s = \frac{\left[\sum_{j=1}^k u_j (d_{2j} - w_{2j}) \right]^2}{\sum_{j=1}^k u_j^2 (v_j)_2}$$

A média de d_{2j} é $w_{2j} = n_{2j} d_j n_j^{-1}$, o que vale a dizer que, se não houver diferença entre duas populações no tempo t_j e v_j é a variância. Com u_j os pesos que especificam os testes. Sob a hipótese nula de que as funções de sobrevivência não diferem, a estatística S tem distribuição qui-quadrado com 1 grau de liberdade para amostras grandes.

Em seguida utiliza-se o teste de Peto, para a comparação função de sobrevivência, como o SPSS só calcula o Breslow e não calcula o teste de Peto. O Software R consegue calcular o teste de Peto e Logrank. De acordo com Colosimo e Giolo (2006), Peto e Peto (1972) e Prentice (1978) sugerem utilizar uma função do peso que depende directamente da experiência passada de sobrevivência observada das duas amostras combinadas.

A função do Peso é uma modificação do Estimador de Kaplan-Meier e é definido de tal forma que seu valor é conhecido antes de falha ocorrer. O estimador modificado da função de sobrevivência é

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

e os pesos utilizados são :

$$u_j = \tilde{S}(t_j - 1) \left(\frac{n_j}{n_j + 1} \right)$$

Portanto, a principal vantagem dos testes Peto-Prentice é que a ponderação é feita relativamente à experiência de sobrevivência anterior. O teste Wilcoxon, em particular, pondera pelo número de indivíduos sob risco que depende da experiência de sobrevivência, assim como de censura.

2.4.2. Modelo Semi-paramétrico de Cox

O modelo de Cox consiste em modelar o logaritmo natural de razão de risco (log razão de risco) através de uma combinação linear de um conjunto de variáveis independentes, de forma muito semelhante à que utilizamos na regressão logística. Assim o modelo de Cox permite a análise de dados provenientes de estudos de tempo de vida em que a resposta é o tempo até à ocorrência de um evento de interesse, ajustado por covariáveis. O modelo de regressão de Cox é o modelo de regressão mais utilizado na análise de dados de sobrevivência, dada a sua flexibilidade e também devido à facilidade de utilização do *software* estatístico.

Júnior Cláudio (2007) refere que o modelo de regressão proposto por Cox tem como principal atractivo a sua simplicidade, abrangência e robustez. Ao contrário dos modelos paramétricos, o modelo de regressão de Cox não exige do utilizador uma escolha *a priori* da distribuição da probabilidade para representar o tempo de sobrevivência. Por ter essa característica, o modelo de regressão de Cox é chamado de modelo semi-paramétrico²⁶. O modelo de **regressão de Cox** faz parte de um grupo de modelos de sobrevivência chamados modelos de risco proporcionais que representam algumas características interessantes que determinam a sua preferência sobre outros modelos.

O modelo de Cox é dado por:

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

em que g é uma função não negativa que deve ser especificada, tal que $g(0) = 1$. Este modelo é composto pelo produto de dois componentes, um não paramétrico e outro paramétrico. O componente não paramétrico, $\lambda_0(t)$, não precisa ser especificado sendo uma função não negativa do tempo. Ele é usualmente chamado de função de base ou basal, pois $\lambda(t) = \lambda_0(t)$ quando $x = 0$. O componente paramétrico é frequentemente usado na seguinte forma multiplicativa:

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

²⁶ Júnior Cláudio (2007), Comportamento reprodutivo Uma análise a partir do grupo ocupacional das mulheres, Brasil, p. 57.

em que β é um vector paramétrico associado às covariáveis e garante que $\lambda(t)$ é não negativa²⁷. A variável dependente é o logaritmo do risco relativo de ocorrer um evento, nos indivíduos em que o evento ainda não ocorreu.

Para ter em conta este facto, a regressão de Cox utiliza o **método de máxima verosimilhança parcial**, Oliveira (2009)²⁸. Portanto, neste método de estimação (maximum likelihood-ML estimation) os parâmetros são calculados ao maximizar a função likelihood, designada por L , para ajustar o modelo de Cox utiliza-se este método.

$$L(\beta) = \prod_{i=1}^k \frac{\exp \{s_i' \beta\}}{\left[\sum_{j \in R(t_i)} \exp \{x_j' \beta\} \right]^{d_i}}$$

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

A estimação dos coeficientes parciais de regressão não é aplicável ao método de máxima verosimilhança, como na regressão logística, porque ao contrário desta, a população em risco não é a mesma em cada dia do estudo devido aos casos censurados e aos que saem da corte por ter ocorrido o evento em estudo.

O modelo estimado é constituído por várias variáveis independentes ou covariáveis tipos binário ou de intervalos, sendo o modelo usualmente referido na literatura dado por:

$$\ln \left(\frac{h(t)}{h_0(t)} \right) = a + b_1 x_1 + b_2 x_2 + \dots + b_n x_n$$

em que x_1, x_2, \dots, x_n são as covariáveis, b_1, b_2, \dots, b_n os respectivos coeficientes parciais de regressão e a a constante de regressão²⁹. Por outro lado, quando a é 0, então se os valores de todas as variáveis independentes forem 0, $h(t)$ terá que ser igual a $h_0(t)$ e o valor da variável dependente será então 0, porque $\ln(1)=0$. Assim o modelo simplifica-se para:

$$\ln \left(\frac{h(t)}{h_0(t)} \right) = b_1 x_1 + b_2 x_2 + \dots + b_n x_n$$

em que $h(t)$ = Taxa de Risco e $h_0(t)$ é risco base no instante t , não especificada.

²⁷ Ibid; 157

²⁸ Oliveira. A. G, Bioestatística Epidemiologia e Investigação, 2009, p. 194.

²⁹ Oliveira. A. G, Bioestatística Epidemiologia e Investigação, 2009, p. 193.

O modelo de Cox é caracterizado pelos coeficientes β 's que medem os efeitos das covariáveis sobre a função de taxa de falha. Segundo Oliveira (2009), uma covariável sobre a razão de risco é constante ao longo do tempo. Assume também que cada covariável aumenta, ou diminui, a taxa de risco numa proporção constante ao longo de todos os seus valores, daí essa relação ser linear numa escala logarítmica. Se experienciarmos a expressão e rearranjarmos os termos, obteremos a expressão que tipicamente descreve o modelo de Cox:

$$h(t) = h_0(t) \exp \{ \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \}^{30}$$

Onde:

$h_0(t)$ = Risco base no instante t

$h(t)$ = Taxa de risco

β_1, \dots, β_p = Parâmetros do modelo

x_1, \dots, x_p = é vector de variáveis explicativas

A fórmula diz que a função *hazard*, que depende de t e \mathbf{x} é composta por dois factores: uma função base $h_0(t)$, que depende de t mas não depende das variáveis explicativas x_i e uma exponencial, que depende apenas das variáveis explicativas x_i , ponderadas pelos respectivos parâmetros β_i . Além disso, a única suposição assumida pelo modelo é a de que, para toda a configuração das variáveis explicativas, as taxas, isto é, as funções *hazard*, são proporcionais.

No caso de uma variável independente binária x , representa-se essa variável por x_0 quando ela toma o valor 0 e por x_1 quando ele toma o valor 1, a razão de risco (HR) entre os dois valores é dada por:

$$HR = \frac{h(t, x_1, \beta)}{h(t, x_0, \beta)}$$

$$HR = \frac{h_0(t) \exp(x_1, \beta)}{h_0(t) \exp(x_0, \beta)}$$

³⁰ Ibid; 194

com esta função pode-se escrever:

$$h(t, x, \beta) = h_0(t) \exp e^{x\beta}$$

Então a expressão é dada por:

$$HR = \frac{h_0(t) \exp (bx_1)}{h_0(t) \exp (bx_0)}$$

Como $h_0(t)$ é assumido igual para todos os indivíduos, podemos eliminar esses termos e ficar com o modelo:

$$HR = \frac{\exp (bx_1)}{\exp (bx_0)}$$

Com $x_0 = 0$, então $\exp (bx_0) = \exp (0) = 1$, a expressão simplifica-se para $HR = \exp (bx_1)$.

Com $x_1 = 1$, o resultado resume-se a $HR = \exp (b)$.

Adicionalmente, sendo a razão de risco uma constante, isso significa que neste modelo se assume que o efeito de um atributo sobre a força de divorciar é constante ao longo do tempo.

O modelo de Cox tem um pressuposto que é necessário que se verifique, chamado pressuposto **de proporcionalidade dos riscos**.

❖ Avaliação dos Riscos Proporcionais

Quando este pressuposto não se verifica, induz ao enviesamento das estimativas dos coeficientes parciais de regressão. Assim, este pressuposto deverá obrigatoriamente ser verificado para cada variável independente no modelo. Para verificar a adequação da suposição de riscos proporcionais, algumas técnicas estatísticas, que incluem análise de resíduos, avaliação gráfica e teste de adequação do ajuste, têm sido propostas.

Colosimo e Giolo (2006), para avaliar a suposição de riscos proporcionais no modelo de Cox, sugerem algumas técnicas gráficas e teste estatísticos que se encontram propostos na literatura. Nesta teoria há três técnicas que são as seguintes: método gráfico descritivo,

método com coeficiente dependente do tempo, método com covariável dependente do tempo³¹.

Os métodos gráficos podem ser divididos basicamente em duas categorias: gerais compostos por gráficos dos resíduos e gráficos Log-log. Utilizam-se ambos para avaliar os riscos proporcionais. O método gráfico descritivo ou **gráfico Log-log** divide por dois ou mais estratos dependendo da categoria da variável. A sua estimação $\hat{\Lambda}_{o,j}(t)$ para cada estrato, o gráfico Log-log é a curva do logaritmo de $\hat{\Lambda}_{o,j}(t)$ versus t, ou log(t) que apresentam diferenças aproximadamente constantes no tempo. A interpretação dos gráficos depende das curvas que se obtêm, por exemplo, curvas não paralelas significam desvios da suposição de risco proporcionais.

Maching, Cheung e Parmar (2006), complementam que o **gráfico Log-log** é muito importante e que o pressuposto é válido para riscos proporcionais e a sua interpretação válida nos coeficientes de regressão do modelo de Cox.. As linhas das curvas dependem dos diferentes valores assumidos pela variável x.

A função de sobrevivência que corresponde o modelo de risco proporcionais, é dada por:

$$S(t; x) = S(t; x_0)^{\exp(\beta x)}$$

onde x e x₀ são valores das variáveis explicativas, em particular x₀ é o valor de variável do grupo de baseline. Os valores de x's são incluídos aqui para enfatizar a sua possibilidade da influência na sobrevivência. O modelo é:

$$\log \{-\log [S(t; x)]\} = \log \{-\log [S(t; x_0)]\} + \beta x$$

Através de métodos dos gráficos, existem três situações que foram consideradas de forma a avaliar o comportamento dos gráficos e dos testes, riscos proporcionais, pequenos afastamentos e grande afastamento. Nessas duas últimas situações os riscos não são proporcionais e definiram-se as condições da seguinte maneira:

- ❖ Riscos Proporcionais: como o próprio nome diz os dados são gerados de maneira que as funções de riscos dos dois grupos são proporcionais;

³¹ Colosimo.E. A e Giolo. S.R, 2006, Análise Sobrevivência Aplicada, 1ª edição, Edgar Blucher, Brasil, pg.166

- ❖ Pequeno afastamento: nesta situação são gerados dados cujos riscos não são proporcionais mas apresentam somente um pequeno grau de afastamento da condição de riscos proporcionais. Além disso, procura-se alguns comportamentos prováveis de acontecerem na prática, como por exemplo: riscos divergentes, convergentes, crescentes e decrescentes;
- ❖ Grande afastamento: foram considerados grandes afastamentos para as funções de risco, isto é, funções de risco que se cruzavam ao longo do tempo³².

Os **resíduos schoenfeld** são definidos para cada falha e não são definidos para censuras. Para definir os resíduos no modelo de Cox, consideramos que o *i*-ésimo indivíduo com vector de covariáveis $x_i = (x_{1i}, x_{2i}, \dots, x_{pi})$ é observado a falha, tem-se para este indivíduo um vector componente r_{iq} , para $q = 1, \dots, p$, que é definido por:

$$r_{iq} = x_{iq} = \frac{\sum_{j \in R(t_i)} x_{jq} \exp \left\{ x'_j \hat{\beta} \right\}}{\sum_{j \in R(t_i)} \exp \left\{ x'_j \hat{\beta} \right\}}$$

Onde r_{iq} = um resíduo apropriado. Como usual para resíduos, $\sum_i r_i = 0$, que permitir que a estrutura de correlação dos resíduos seja considerada, uma forma padronizada dos resíduos de Schoenfeld (Scaled Schoenfeld residual) e frequentemente usada e é definida por:

$$S_i^* = \left[I(\hat{\beta}) \right]^{-1} \times r_i,$$

com $I(\hat{\beta})$ a matriz de informação observada. O modelo de expressão é dado por:

$$\lambda(t) = \lambda_0(t) \exp \{ x' \beta(t) \}$$

Com a restrição de que $\beta(t) = \beta$, como uma alternativa de representar o modelo de Cox. Observa-se que $\beta(t) = \beta$ implica na proporcionalidade do risco. Quando $\beta(t)$ não é constante, o impacto de uma ou mais covariáveis no risco pode variar com o tempo.

³² Cardoso e Colosimo (2003) Avaliação da suposição de risco proporcionais para o modelo de Cox, 48ª Reunião da RBRAS e 10ª Seagro, p.3

O resíduo Schoenfeld pode ser calculado para cada variável explicativa x_k e cada observação não censurada j , cujo tempo de falha t_j é dado por:

$$SCH_{kj} = x_{kj} - E(x_{kj} | R_j)$$

onde R_j é o conjunto de observações em risco de falha no tempo falha j enquanto $E(x_{kj} | R_j)$ é o valor esperado de x_k para j estimado com base no log HR de um modelo de Cox.

Nos resíduos padronizados Schoenfeld existe duas maneiras para se resolver, os gráficos dos resíduos e teste de Honell's ou coeficiente de correlação de Pearson (ρ) entre os resíduos padranizados de Schoenfeld e $g(t)$, para cada covariável, é uma dessas medidas para avaliar o risco proporcional. Valores de ρ próximos de zero mostram não haver evidências para a rejeição da suposição de riscos proporcionais.

A hipótese geral (ou global) de proporcionalidade dos riscos para todas as covariáveis no modelo assumindo $g_q(t) = g(t)$, pode se realizado pelo teste de estatística:

$$T = \frac{\left(\bar{g} - \bar{g} \right)' S^* I S^* \left(\bar{g} - \bar{g} \right)}{d \sum_k \left(\bar{g}_k - \bar{g} \right)^2}$$

em que I é a matriz de informação observada, d é o número de falhas e $S^* = dR\Gamma^1$, sendo R a matriz $d \times p$ dos resíduos Schoenfeld não padronizados. Sob a hipótese nula de proporcionalidade dos riscos aproximadamente distribuição qui-quadrado com p graus liberdade. Valores de $T > x_{p,1-\alpha}^2$ mostram-se evidências contra a suposição de riscos proporcionais

De acordo com Giolo e Colosimo (2006) os **resíduos martingal** compostos por ponto atípicos e forma de funcionamento das covariáveis. Assim, quando os dados apresentam censura direita e todas covariáveis são fixadas no início do estudo ou não forem dependentes do tempo, os resíduos martingal para o modelo de Cox são definidos por:

$$\hat{m}_i = \delta_i - \Lambda_0 \hat{\Lambda}(t_i) \exp \left\{ \sum_{k=1}^p x_{ip} \hat{\beta}_k \right\} = \delta_i - \hat{e}_i, \dots, i = 1, \dots, n.$$

Esses resíduos são frequentemente usados para verificar a presença de pontos atípicos (outlier), bem como para verificar a forma funcional das covariáveis, isto é, se estas

devem ser usadas no modelo como $\log(x_i)$, x_i^2 e assim por diante, em vez de x_i , ou mesmo categorizadas³³. Os resíduos martingal, são usados para determinar a forma funcional das covariáveis a serem incluídas no modelo e, ocasionalmente, úteis para avaliar a adequação do ajuste. Outro resíduo usado, em geral, com o propósito de detectar pontos atípicos (Outliers), é o **resíduo deviance**. Esses resíduos são definidos por:

$$\hat{d}_i = \sin al \left(\hat{m}_i \right) \left[- 2 \left(\hat{m}_i + \delta_i \log \left(\delta_i - \hat{m}_i \right) \right) \right]^{\frac{1}{2}}$$

O gráfico \hat{d}_i versus o predictor linear $\sum_{l=1}^p x_{il} \beta_l, i=1, \dots, n$ é utilizado, nesse caso, para avaliar a presença de dados atípicos.

Os resíduos **Cox-Snell** são úteis para avaliar a qualidade do ajuste para o modelo global. Esses resíduos são normalmente usados para validar um determinado modelo que foi ajustado aos dados considerando que se trata de um modelo de *hazards* proporcionais. Começemos por considerar os resíduos de Cox-Snell, definidos por,

$$\hat{r} = \hat{H} (T_i | Z_i),$$

onde \hat{H} é a estimativa da função *hazard cumulativa* para o modelo ajustado³⁴.

Segundo Turkman e Silva (2000), um indicador da **influência** da *i*-ésima observação (y_i, z_i), no vector estimado $\hat{\beta}$, pode ser calculado pela diferença $\hat{\beta} - \hat{\beta}_{(i)}$, onde $\hat{\beta}$ e $\hat{\beta}_{(i)}$ representam, respectivamente, as estimativas de máxima verosimilhança do vector parâmetro β obtidas da amostra sem observação (y_i, z_i) e da amostra com todas as observações. Se $\hat{\beta}_{(i)}$ for substancialmente diferente de $\hat{\beta}$, então a observação (y_i, z_i) pode ser considerada influente³⁵. A medida da influencia do **Cook** é dada por:

$$D_i = \frac{\left(\hat{\beta} - \hat{\beta}_{(i)} \right)^T \left(Z^T W Z \right) \left(\hat{\beta} - \hat{\beta}_{(i)} \right)}{p \hat{\phi}}$$

De acordo com Giolo e Colosimo (2006), outro uso importante dos resíduos é auxiliar na avaliação da influência (impacto) de cada observação no ajuste de um modelo. O

³³ Giolo. S.R e Colosimo.E. A, 2006, *Análise Sobrevida Aplicada*, 1ª edição, Edgar Blucher, Brasil, pg.173

³⁴ Dias Inês, (Modelos de Tempo de Vida Acelerado em Análise sobrevida, Tese, U. Lisboa, 2004

³⁵Turkman e Silva (2000), *Modelos Lineares Generalizado*, Trabalho Teoria e Pratica, Lisboa

modelo global de efeito das observações pode ser obtido da seguinte forma:

$$\Delta\beta_i = \left(\hat{\beta} - \hat{\beta}_{(i)} \right)' \left[I \left(\hat{\beta} \right) \right]^{-1} \left(\hat{\beta} - \hat{\beta}_{(i)} \right), i = 1, \dots, n,$$

Com $\Delta\beta$ a mudança no vector de coeficientes estimados obtida pela remoção, uma de cada vez, das observações. O gráfico desses resíduos versus i pode ser útil na detecção de observações influentes. Maching, Cheung e Parmar (2006), referem que quando não se verifica a proporcionalidade existem várias opções a fazer:

- ❖ Análise por estratos da variável exposição; ou seja, não criar nenhum modelo e fazer curvas de Kaplan-Meier para cada grupo separadamente;
- ❖ Iniciar a análise a determinado dia, e utilizar o modelo de Cox nos que sobreviveram após esse dia;
- ❖ Estimar um modelo de Cox até determinado dia e outro após esse dia, de modo a obter duas diferentes estimativas de hazard ratio (HR), uma para cada um dos períodos;
- ❖ Estimar um modelo de Cox modificado que inclua uma variável de tempo dependente que meça a interação da exposição com o tempo. Este modelo é designado por modelo de Cox estratificado ou estendido (extended Cox model)³⁶

2.4.3. Modelos Probabilísticos – Análise de Sobrevivência

Apesar de existirem técnicas não paramétricas que tratam de dados sobre análise de sobrevivência, deve-se começar por tentar ajustar os dados ao modelo paramétrico mais apropriado. Esses modelos foram: a distribuição exponencial, a distribuição de Weibull, a log-normal e log-logística.

❖ Distribuição Exponencial

Em termos matemáticos, a **distribuição exponencial** é um dos modelos probabilísticos mais simples usados para descrever o tempo de falha. Ela tem sido extensivamente usada para descrever o tempo de vida de certos produtos e materiais e tem descrito adequadamente o tempo de vida de óleos isolantes e dielétricos, entre outros. Esta

³⁶ Machin, Cheung, Parmar (2006), Survival Analysis A Practical Approach, 2nd Edition, Enland, Jhon wiley & Sons, Ltd, p. 149.

distribuição apresenta um único parâmetro e é a única que se caracteriza por ter uma função de taxa de falha (ou de risco) constante³⁷.

A variável aleatória tempo de falha, T, tem distribuição exponencial, com parâmetro $\alpha > 0$, a função densidade de probabilidade, $f(t)$, é dada por:

$$f(t) = \frac{1}{\alpha} \exp\left\{-\left(\frac{t}{\alpha}\right)\right\}, t \geq 0$$

Onde o parâmetro $\alpha > 0$ representa é o tempo médio de vida. As funções de sobrevivência, $S(t)$ e taxa de falha $\lambda(t)$ são:

$$\hat{S}(t) = \exp\left\{-\left(\frac{t}{\alpha}\right)\right\}, t \geq 0 \quad \text{e} \quad \lambda(t) = \frac{1}{\alpha} \text{ para } t \geq 0$$

Outras características de interesse são a média, a variância e os percentis. O percentil 100p% corresponde ao tempo em que 100 % dos que falharem no casamento, o p-ésimo quantil é $\alpha \log(1 - p)$. A média da distribuição exponencial é α , e a variância α^2 . Os percentis são importantes quando queremos obter informações, por exemplo, a respeito da ocorrência de eventos prematuros.

❖ Distribuição Weibull

A **distribuição de Weibull** é uma função de verosimilhança para uma amostra de tempo de vida. A distribuição Weibull foi uma proposta originalmente por Weibull em 1939 e sua ampla aplicabilidade foi também discutida por este mesmo autor (Weibull, 1951, 1954). Desde então, a mesma vem sendo frequentemente usada em estudos biomédicos e industriais. A sua popularidade em aplicações práticas deve-se ao facto dela apresentar uma grande variedade de formas, todas com uma propriedade básica: a sua função de taxa de falhas pode assumir várias formas: crescente, decrescente ou constante³⁸.

Uma variável T segue distribuição de Weibull a sua função densidade de probabilidade, a sua função de sobrevivência e a função taxa de risco são dadas,

$$f(t) = \left(\frac{\beta}{\alpha^\beta}\right) t^{\beta-1} \exp\left[-\left(\frac{t}{\alpha}\right)^\beta\right], t \geq 0$$

³⁷ Colosimo.E. A e Giolo. S.R, 2006, Análise Sobrevivência Aplicada, 1ª edição, Edgar Blucher, Brasil, pg.70

³⁸ Ibid:73

onde de parâmetros β e α , isto é, $W(\beta, \alpha)$, sendo $\beta > 0$ o parâmetro de escala e $\alpha > 0$ o parâmetro de forma. Para $t \geq 0$. É o modelo de Weibull para análise sobrevivência e taxa de risco dada por:

$$\hat{S}(t)_W = \exp \left\{ - \left(\frac{t}{\alpha} \right)^\gamma \right\}$$

$$\lambda_t = \frac{\gamma}{\alpha^\gamma} t^{\gamma-1}$$

Para $t \geq 0$ e $\gamma > 0$. O parâmetro α é o parâmetro de escala e γ o de forma, ambos são positivos (Colosimo; Giolo, 2006). Se $\gamma = 1$ tem-se a taxa de risco constante, esse é o caso da distribuição exponencial, que é um caso particular da Weibull. Se $\gamma > 1$ a taxa de risco é estritamente crescente e no caso $\gamma < 1$ a taxa é estritamente decrescente.

A média e a variância para a distribuição de Weibull apresentam as seguintes expressões:

$$E(T) = \alpha \Gamma \left(1 + \left(\frac{1}{\gamma} \right) \right),$$

$$Var(T) = \alpha^2 \left[\Gamma \left(1 + \left(\frac{2}{\gamma} \right) \right) - \Gamma \left(1 + \left(\frac{1}{\gamma} \right) \right)^2 \right],$$

Onde $\Gamma(k)$ é a função gama dada por $\Gamma(k) = \int_0^k x^k \exp\{-x\} dx$.

A distribuição Weibull é uma distribuição flexível, uma vez que permite uma função de risco decrescente ($\alpha < 1$), uma função de risco crescente ($\alpha > 1$), e o risco constante quando $\alpha = 1$.

❖ Distribuição Log-normal

A **distribuição log-normal** é muito utilizada para caracterizar tempos de vida de produtos, indivíduos e bastante utilizada para descrever situações clínicas, como o tempo de vida de pacientes com Leucemia³⁹. A distribuição log-normal, de uma forma mais simples, pode ser definida como a distribuição cujo logaritmo da variável aleatória T , segue a distribuição normal. Deste modo, uma variável que segue uma distribuição com parâmetros μ e σ com a função densidade de probabilidade dada por

³⁹ Ibid: 75

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(\log t - \mu)^2\right], t \geq 0$$

Em que μ é média do logaritmo do tempo de falha, assim como σ é o desvio padrão e que:

❖ Média: $E[T] = \exp\left(\mu + \frac{\sigma^2}{2}\right)$.

❖ Variância: $Var[t] = \exp(2\mu + \sigma^2) [\exp(\sigma^2) - 1]$

Função de sobrevivência e taxa de risco de uma distribuição Log-normal é dada por:

$$\hat{S}(t)_{LN} = \Phi\left(\frac{-\log(t) + \mu}{\sigma}\right)$$

$$\lambda(t)_{LN} = \left(\frac{f(t)}{S(t)}\right)$$

Em que $\Phi(\cdot)$ é a função de distribuição acumulada de uma normal padrão. As funções de taxa de falha não são monótonas: crescem, atingem um valor máximo e depois decrescem.

❖ Distribuição Log-logística

Uma distribuição que em muitas situações práticas, tem se apresentado com uma alternativa à de Weibull e à Log-normal é a log-logística. A sua função de distribuição tem uma expressão aberta e a sua função de risco aumenta inicialmente e depois diminui. A função densidade e as funções de sobrevivência com os parâmetros α e λ , ambos estritamente positivo, são dados dada por:

$$f(y) = \frac{\lambda \alpha y^{\alpha-1}}{(1 + \lambda y^\alpha)^2}$$

$$S(y) = \frac{1}{1 + \lambda y^\alpha}$$

$$\hat{S}(t)_{LL} = \frac{1}{1 + \exp\left\{\frac{\log(t) - \mu}{\sigma}\right\}}$$

$$\lambda(t)_{LL} = \frac{1}{\sigma} \exp\left\{\frac{y - \mu}{\sigma}\right\} \left(1 + \exp\left\{\frac{y - \mu}{\sigma}\right\}\right) - 1$$

2.4.4. Estimação dos Parâmetros do Modelo

Na análise de sobrevivência precisamos de estimar os parâmetros dos modelos. Entretanto, em cada estudo envolvendo tempos de falha, os parâmetros devem ser estimados a partir das observações para que o modelo fique determinado e assim ser possível responder as perguntas de interesse.

Um dos métodos para estimar os parâmetros do modelo é o **método de máxima verosimilhança** que surge como uma opção apropriada para estes tipos de dados. Ele incorpora as censuras, é relativamente simples de ser entendido e possui propriedades ótimas para grandes amostras Giolo e Colosimo (2006)⁴⁰. A função de uma verosimilhança para um parâmetro genérico θ de uma população é dada por:

$$L(\theta) = \prod_{i=1}^n f(t_i; \theta).$$

O método de máxima verosimilhança somente pode ser aplicado após ter sido definido um modelo probabilístico adequado para os dados.

2.4.5. Métodos dos Gráficos para Escolher o Melhor Modelo

❖ Comparação das curvas de sobrevivência dos modelos com as curvas de sobrevivência de KM

Conforme Colosimo e Giolo (2006), o primeiro método gráfico para escolher o melhor modelo a ser usado consiste na comparação da função sobrevivência do modelo proposto com estimador de Kaplan-Meier $\hat{S}(t)$, nesse método ajusta-se o modelo proposto (exponencial, Weibull, Log-normal) e estima-se as funções de sobrevivência a partir das estimativas dos parâmetros de cada modelo. Obtém-se também a estimativa de Kaplan-Meier para a função de sobrevivência e, em seguida comparam-se graficamente as funções sobrevivências de Kaplan-Meier e do modelo proposto. O modelo que melhor se ajusta aos dados é aquele que mais se aproximada curva de Kaplan-Meier.

Alternativamente pode ser colocar no mesmo gráfico a estimada por Kaplan-Meier versus o tempo e a estimada pelo modelo proposto versus o tempo. Dessa maneira, o modelo

⁴⁰ Ibid; 83-84

cuja curva mais se aproximar da curva Kaplan-Meier indicará modelo que melhor se ajusta aos dados (Colosimo & Giolo, 2006).

O segundo método gráfico, apresentado por Colosimo e Giolo (2006) consiste na linearização da função de sobrevivência. Ao linearizar a função tem-se como o resultado uma recta, caso o modelo seja adequado.

❖ Métodos dos gráficos

A escolha do modelo paramétrico na análise sobrevivência na forma mais simples e eficiente de seleccionar o ser usado para um conjunto de dados é por meio de **técnicas gráficas**. Para ajustar os modelos propostos ao conjunto de dados (por exemplo, os modelos log normal, Weibull) e a partir das estimativas dos parâmetros de cada modelo, estimam-se as respectivas funções de sobrevivência, representadas pelos modelos log-normal e de weibull por $\hat{S}_{ln}(t)$ e $\hat{S}_w(t)$ respectivamente e obtém-se, também, a estimativa de Kaplan-Meier para a função de sobrevivência $\left(\hat{S}(t)\right)$ ⁴¹. Nos modelos paramétricos para escolher qualquer um deste modelo, inicialmente utiliza-se o método de gráfico, e este método consiste na linearização da função de sobrevivência tendo como ideia básica a construção de gráficos que sejam aproximadamente lineares, caso o modelo proposto seja apropriado.

a. Linearização no modelo Exponencial

A linearização do modelo exponencial, a função de sobrevivência, dada por:

$$S(t) = \exp \left\{ - \left(\frac{t}{\alpha} \right) \right\},$$

Assim,

$$- \log [S(t)] = \frac{t}{\alpha} = \left(\frac{1}{\alpha} \right),$$

O que mostra que $-\log[S(t)]$ é uma função linear de t . Logo, o gráfico de $-\log[\hat{S}(t)]$ versus t deve ser aproximadamente linear, passando pela origem, se o modelo exponencial for apropriado, $\hat{S}(t)$ é o estimador de Kaplan-Meier.

⁴¹ Ibid;96

b. Linearização no modelo Weibull

A função de sobrevivência para o modelo de Weibull de parâmetros (γ, α) é, como visto anteriormente dada por:

$$S(t) = \exp \left\{ - \left(\frac{t}{\alpha} \right)^\gamma \right\}, t \geq 0$$

Desse modo,

$$- \log [S(t)] = \left(\frac{t}{\alpha} \right)^\gamma$$

$$- \log [- \log [S(t)]] = -\gamma \log [\alpha] + \gamma \log (t)$$

c. Linearização no modelo Log-normal

O que mostra que $\log [- \log [S(t)]]$ é uma função linear de $\log (t)$. Por tanto, o gráfico de $\log [- \log [S(t)]]$ versus $\log (t)$, sendo $\hat{s}(t)$ é o estimador de Kaplan-Meier, deve ser aproximadamente linear, se o modelo de Weibull for apropriado. Se além de linear, o gráfico passar pela origem e tiver inclinação igual a 1, é uma indicação a favor do modelo exponencial.

A função de sobrevivência para o modelo Log-normal dada por:

$$S(t) = \Phi \left(\frac{- \log (t) + \mu}{\sigma} \right)$$

Pode ser linearizada, e apresenta a seguinte forma:

$$\Phi^{-1} (S(t)) = \frac{- \log t + \mu}{\sigma}$$

Em que $\Phi^{-1} (\cdot)$ são os percentis da distribuição normal padrão. Isto significa que o gráfico de $\Phi^{-1} (\hat{s}(t))$ versus $\log (t)$ deve ser aproximadamente linear, com intercepto μ/σ e inclinação $-1/\sigma$, se o modelo log-normal for apropriado.

Na seguinte fazemos comparações dos modelos para seleccionar qual é melhor modelo ou modelo mais adequado. Outra forma de discriminar modelos é por meio de teste

de hipótese⁴². Portanto para testar os testes de hipótese utiliza-se a **estatística de razão verosimilhanças**, mas apenas em modelos encaixados (Cox e Hinkley, 1974). Isto significa que deve ser identificado um modelo generalizado tal que os modelos de interesse sejam casos particulares.

O teste realizado a partir dos seguintes:

- a. Modelo generalizado e obtenção do valor do logaritmo de sua função de verosimilhança $\left(\log L\left(\hat{\theta}_G\right) \right)$
- b. Modelo de interesse e obtenção do valor de logaritmo de sua função de verosimilhança $\left(\log L\left(\hat{\theta}_M\right) \right)$.

A partir desses valores é possível calcular a estatística da razão de verosimilhanças, isto é:

$$TRV = -2 \log \left[\frac{L\left(\hat{\theta}_M\right)}{L\left(\hat{\theta}_G\right)} \right] = 2 \left[L\left(\hat{\theta}_G\right) - L\left(\hat{\theta}_M\right) \right],$$

que, sob o H_0 , tem aproximadamente uma distribuição qui-quadrado com graus liberdade igual a diferença do número de parâmetros ($\hat{\theta}_G$ e $\hat{\theta}_M$) dos modelos sendo comparados.

Pode também usar-se o critério de **Akaike (AIC)**. Assim, vamos indicar um outro método, proposto por Akaike (1974), para seleccionar o modelo paramétrico que melhor se adapta aos dados, que tem em conta não só a verosimilhança como o número de parâmetros do modelo. Pretende-se encontrar um modelo que se ajuste bem aos dados e que contenha o menor número de parâmetros possível. Para modelos paramétricos tem-se

$$AIC = -2 \ln \hat{L} + 2(p+1+k)$$

onde p é o número de covariáveis no modelo e o valor de k é escolhido de acordo com o número de parâmetros de cada modelo. Considera-se $k = 0$ para o modelo exponencial, $k = 1$ para os modelos Weibull, log-logístico e log-normal e $k = 2$ para o modelo gama generalizado⁴³. A partir de critério AIC é possível escolher o modelo mais adequado para os dados sendo preferível o modelo que tem menor AIC.

⁴² Ibid; 100

⁴³ Dias, Inês, (Modelos de Tempo de Vida Acelerado em Análise sobrevivência, Tese, U. Lisboa, 2004

2.5. Modelos de Fragilidade

O modelo fragilidade num contexto univariado é uma medida da heterogeneidade dos indivíduos, por outro lado permite-se avaliar o efeito de covariáveis que por algum motivo não foram observadas na ocasião do experimento, e desse modo, não foram incluídas no modelo⁴⁴. A análise de fragilidade como uma possibilidade de dependência e heterogeneidade não observadas por outros métodos. O modelo de fragilidade pode ser visto como um modelo de efeitos aleatórios para dados de eventos recorrentes no tempo, onde o efeito aleatório tem um efeito multiplicativo na função de risco base. A heterogeneidade dos indivíduos afecta, portanto, os tempos de observados e para considera-la um efeito aleatório (Colosimo & Giolo, (2006). um modo conveniente de se explicar a heterogeneidade das variáveis é obtido através do modelo proposto da função de risco e sobrevivência não condicionais à fragilidade, com parâmetros de interesse e covariáveis.

$$\lambda_i(t) = z_i \lambda_0(t) \exp \{X_i' \beta\}$$

Onde:

β = o vector de parâmetros desconhecidas associados às covariáveis X_i ,

$\lambda_i(t)$ = a função de risco de base não especificada

z_1, \dots, z_n = os valores das fragilidades, assumidas serem uma amostra provenientes variáveis aleatórias Z_i independentes e identicamente distribuídas com distribuição de probabilidade conhecida de média 1 e com variância desconhecida. Note-se que as variáveis de fragilidade Z_i não variam com o tempo, o que pode representar uma limitação deste modelo.

Nesta análise da fragilidade apenas fazer para averiguar o modelo Cox, ou seja tenta-se o modelo mais avalia ou melhorar a análise, portanto ver se existe algumas heterogeneidades entre as variáveis.

⁴⁴ Colosimo E. A, Giolo S. R, Análise sobrevivência aplicada, 1ª edição, Edgar Blucher, Brasil, pg. 282-285.

CAPÍTULO III

Caracterização dos Imigrantes do Distrito de Évora

3.1. Descrição das Variáveis

Os imigrantes que responderam ao questionário do SEF foram 354 indivíduos. Os respondentes foram questionados relativamente a um conjunto de dados que permite caracterizar as suas situações. Os dados recolhidos através do questionário foram tratados estatisticamente.

As variáveis que foram analisadas para caracterizar os imigrantes no Distrito de Évora foram: nacionalidade dos imigrantes, idade, naturalidade, sexo, estado civil, designação da morada, motivo de entrada, profissão e nível da escolaridade.

As variáveis que foram analisadas para caracterizar os casamentos dos imigrantes foram: números dos casamentos dos imigrantes, nacionalidade do cônjuge, idade, sexo, estado civil, residência antes e depois de casamento, naturalidade do cônjuge, celebração do casamento, motivo de entrada, nível de escolaridade, tempo (duração do casamento), número de filhos do casamento, número de filhos de outro casamento, profissão do cônjuge, nacionalidade dos imigrantes (países União Europeia ou não), status (ocorrência de divórcio ou não). As variáveis não numéricas foram codificadas de acordo com as tabelas apresentadas em anexo.

3.2. Análise Exploratória de Dados

Através das respostas aos questionários procurámos, em primeiro lugar, identificar as seguintes características pessoais dos imigrantes.

❖ Sexo

Através da tabela e do gráfico que se seguem, verificámos que a distribuição dos respondentes por sexo é diferente, com vantagem para o grupo feminino que representa cerca de dois terços dos respondentes. No total os estrangeiros com estatuto de comunitário são 242 indivíduos do sexo feminino e 112 do sexo masculino. Podemos dizer que os imigrantes comunitários inscritos são maioritariamente do sexo feminino, o que obviamente se reflecte também na representação das nacionalidades (Ver Tabela 1 e Figura 1).

Tabela 1. Distribuição dos imigrantes por sexo

Sexo	Frequência	Percentagem
Feminino	242	68,4
Masculino	112	31,6

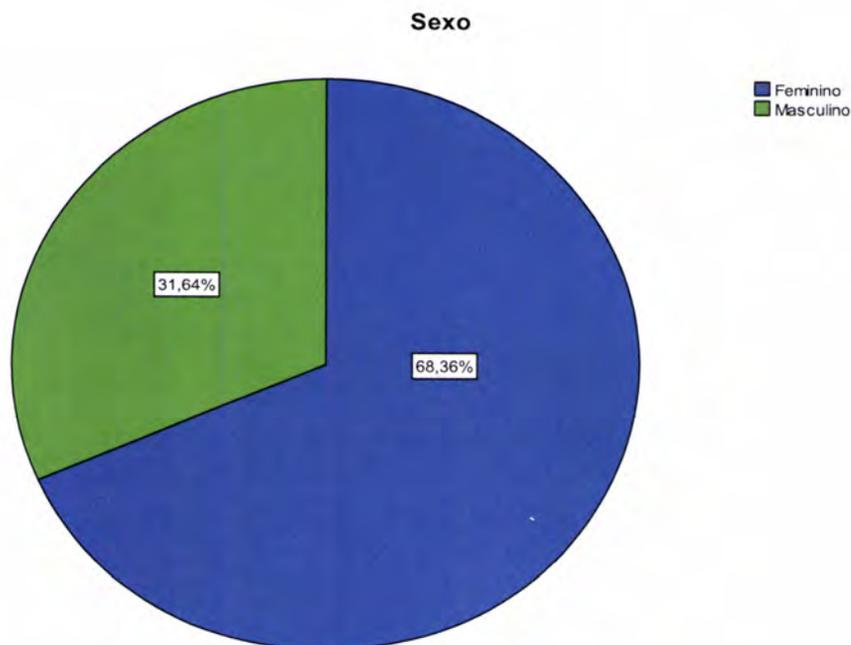


Figura 1. Gráfico circular por sexo

❖ Nacionalidade

Considerando a totalidade dos imigrantes que fazem parte dos comunitários, a mesma é composta por 39 nacionalidades no período entre 2006 e 2009 em Évora. Verifica-se que estes imigrantes são na sua maioria de nacionalidade brasileira, com uma percentagem superior a 46,6% dos 354 inscritos legalmente nos Serviço de Estrangeiros e Fronteiras, de Évora. A seguir surge a nacionalidade Holandesa, mas apenas com 5,9%, e vai diminuindo sucessivamente para outras nacionalidades até atingir 0,3% para algumas nacionalidades, como, por exemplo, Costa Rica (Tabela 2 e Figura 2).

Tabela 2. Distribuição de imigrantes por nacionalidade

Nacionalidade dos imigrantes		
Nacionalidade	Frequência	Porcentagem
Brasileira	165	46,6
Romena	32	9
Holandesa	21	5,9
Alemã	19	5,4
Espanhola	17	4,8
Ucraniana	11	3,1
Italiana	11	3,1
Cabo-verdiana	8	2,3
Búlgara	8	2,3
Lituana	7	2
Moçambicanos	4	1,1
Francesa	4	1,1
Americanos	3	0,8
São-tomense	3	0,8
Angolana	3	0,8
Belga	3	0,8
Russa	3	0,8
Húngara	3	0,8
Costa Marfim	2	0,6
Guineense	2	0,6
Moldova	2	0,6
Chinesa	2	0,6
Venezuelana	2	0,6
Marroquina	2	0,6
Rep. Checa	2	0,6
Britânica	2	0,6
Boliviana	1	0,3
Peruana	1	0,3
Colombiana	1	0,3
Nova Zelândia	1	0,3
Vietnamita	1	0,3
Filipina	1	0,3
Argentina	1	0,3
Suíça	1	0,3
Finlandesa	1	0,3
Polaco	1	0,3
Irlandesa	1	0,3
Estoniana	1	0,3
Costa Rica	1	0,3
Total	354	100

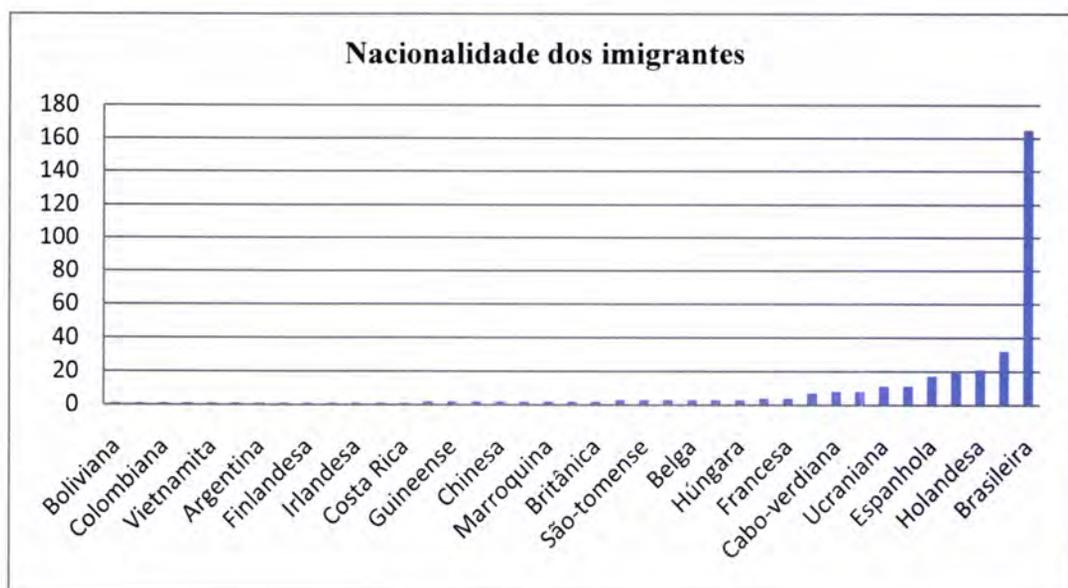


Figura 2. Diagrama de barras por nacionalidade dos imigrantes

❖ Nacionalidades com codificação por classe

Como observámos para algumas nacionalidades uma frequência muito pequena agrupámos nacionalidades por classes de nacionalidade (Para a Tabela 3 relativamente à Tabela 2 e Figura 3).

Como afirmámos anteriormente, foram consideradas quatro classes de nacionalidades principais nas opções de resposta à questão relativa à nacionalidade dos imigrantes: Brasileira, Africanos, Europeus e as restantes nacionalidades. Como se pode verificar (Tabela 3), a maioria dos nossos imigrantes comunitários é de origem Brasileira que representa 46,6%, seguidos de perto pelos da União Europeia, e em menor número pelos Africanos e pelos restantes. Das opções apresentadas, as “restantes nacionalidades” é a classe de nacionalidade menos representada nesta população, havendo países restantes só com um imigrante em Évora, por exemplo, Colômbia e Vietname. A distribuição dos respondentes por classes de nacionalidades é seguinte:

Tabela 3. Distribuição de imigrantes por classes de nacionalidades

Codificação Nacionalidade dos Imigrantes		
	Frequência	Percentagem
Brasil	165	46,6
União Europeia	111	31,4
Restantes	56	15,8
África	22	6,2
Total	354	100

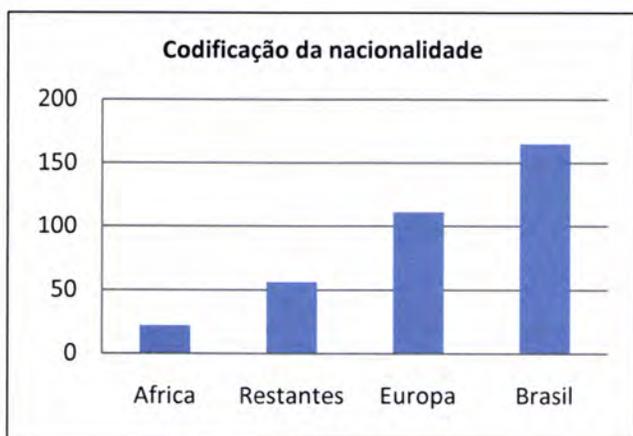


Figura 3. Diagrama de barra por codificação de nacionalidade

A seguir mostra-se as entradas de imigrantes em cada ano por classes de nacionalidade.

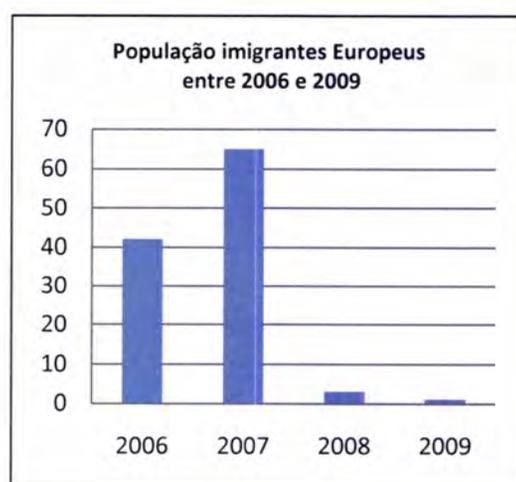
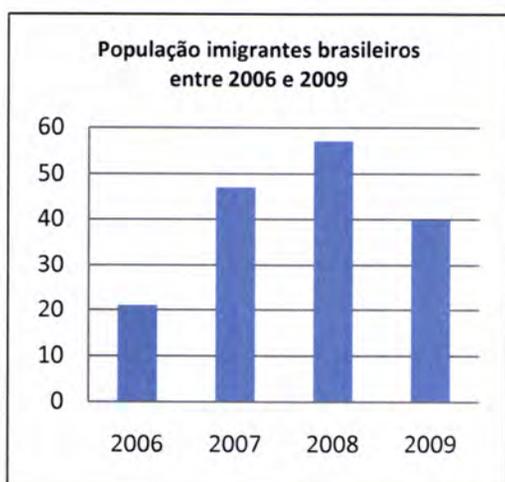


Figura 4 e 5: Diagramas de barras por populações imigrantes Europeus e Brasileiros de 2006 a 2009

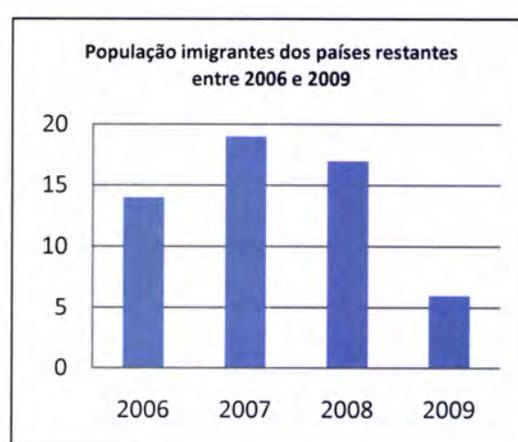
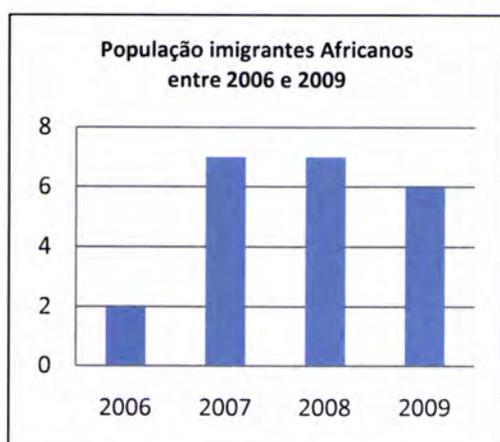


Figura 6 e 7: Diagramas de barras por populações imigrantes Africanos e “Restantes Países” de 2006 a 2009

A entrada de brasileiros foi crescente em 2006, 2007 e 2008, verificando-se uma descida de 2008 para 2009. Os imigrantes que fazem parte da União Europeia que entraram em Évora cresceu entre 2006 e 2007, mas nos anos seguintes decresceu bastante; isto pode dever-se ao facto desses imigrantes passarem a pedir os certificados na Câmara Municipal, e não passarem pelo SEF. De 2006 para 2007 a entrada de africanos subiu bastante, em 2007 e 2008 os africanos mantêm igual o número de entradas, registando-se uma ligeira descida em 2009. Para as restantes nacionalidades verifica-se um crescimento entre 2006 e 2007, mas registou-se uma queda acentuada em 2009 (Figuras 4, 5, 6, 7).

❖ Nível de Escolaridade

O perfil por nível de escolaridade caracteriza-se pelas baixas qualificações literárias por parte da generalidade dos 354 imigrantes comunitários inscritos no SEF de Évora. Através da tabela e da figura que se seguem, podemos verificar que a maioria dos imigrantes comunitários tem nível escolaridade até ao secundário. No entanto, 45,7 % dos indivíduos tem curso secundário; ao nível de escolaridade básica corresponde 23,6%, quase o mesmo que o nível escolaridade médio ou superior (23,3%). Na categoria técnico profissional corresponde apenas 5,6% dos imigrantes comunitários. Os respondentes que não têm grau de escolaridade representam 1,9%. Esta informação pode-se observar na Tabela 4 e Figura 8.

Tabela 4. Distribuição dos imigrantes por nível escolaridade

Nível de escolaridade		
	Frequência	Percentagem
Secundário	147	45,7
Básico	76	23,6
Superior	75	23,3
Técnico profissional	18	5,6
Nenhum	6	1,9
Total	322	100
NA	32	
Total	354	

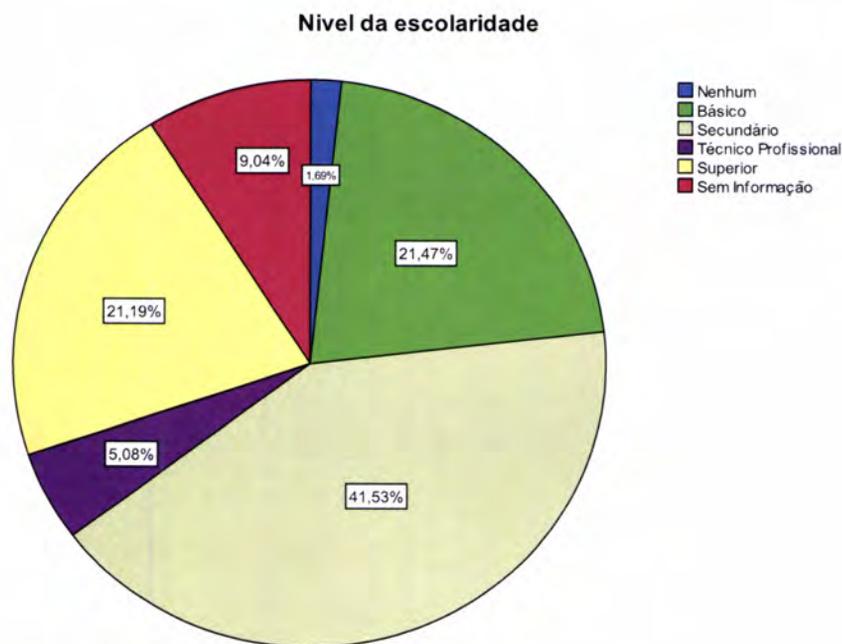


Figura 8. Gráfico circular por nível de escolaridade

❖ Idade dos imigrantes

A idade é uma característica importante para podermos caracterizar os imigrantes comunitários que entram em Portugal, em Évora. São pessoas com uma idade média de 35 anos.

Pode-se verificar as idades na Tabela 5 e Figura 9 e 10. A idade dos imigrantes varia entre os 3 e os 87 anos, sendo que a moda são os 31 anos e a mediana das idades se situa nos 33 anos.

Tabela 5. Distribuição dos imigrantes por idade

Estatística	
Idade dos imigrantes	
N	354
Média	34,65
Mediana	33,00
Moda	31,00
Desvio Padrão	14,82
Variância	219,69
Mínimo	3
Máximo	87
Percentis 25	26
50	33
75	44

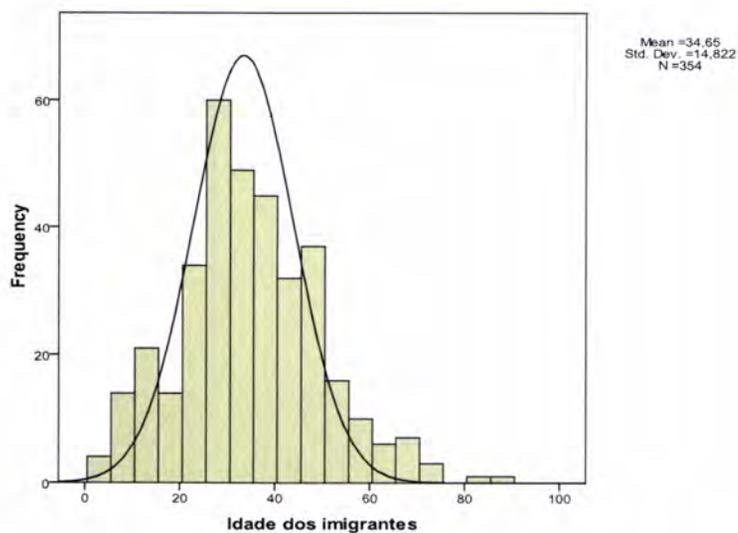


Figura 9. Histograma e curva normal para Idade dos imigrantes

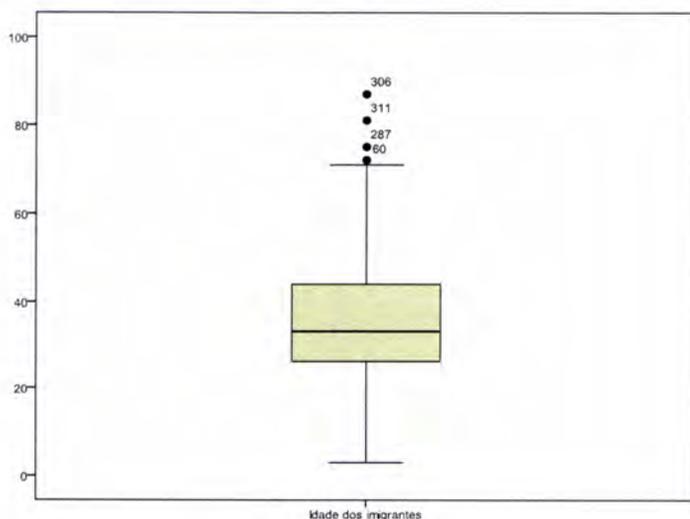


Figura 10. Caixa de bigodes por idade

A caixa de bigodes dá informação sobre a localização central, a mediana que varia entre idade 20 e 40 ou seja o valor de mediana é 33. A média é 34,65 o que pode afectar os outliers, outras localização é 1º e 3º quartis e mínimo e máximo. Em seguida a dispersão: amplitude e distância inter-quartil, Assimetria: posição relativa da mediana na caixa, comprimento dos bigodes. Por outro lado, observa-se na caixa de bigodes existe também os outliers. Esses outliers são indivíduos que têm idade 72, 75, 86 e 81 anos. Finalmente, na Figura 11 e 12 mostram-se assimetria negativa e a distribuição mais ponteguada ou Leptocurtica.

❖ Estado Civil

A maioria dos imigrantes comunitários e os imigrantes que têm relação com nacionais da União Europeia) no Distrito de Évora são oficialmente casados (65,1%), seguida pelos indivíduos solteiros (30,3%); os divorciados são 2,9%, os viúvos são 1,4% e, por fim, os separados são apenas 0,3%. Veja-se esta informação na Tabela 6 e Figura 10.

Tabela 6. Distribuição dos imigrantes por estado civil

Estado Civil		
	Frequência	Percentagem
Casados	226	65,1
Solteiros	105	30,3
Divorciadas	10	2,9
Viúvas	5	1,4
Separadas	1	0,3
Total	347	100
NA	7	
Total	354	

Em seguida mostra-se o diagrama de barras para a população imigrante por estado civil.

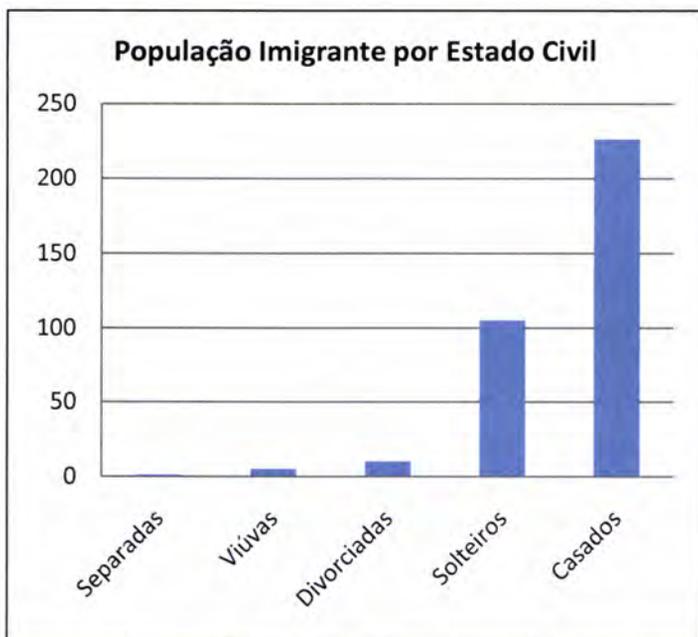


Figura 11. Diagrama de barras por estado civil

❖ Motivo de entrada

Os imigrantes têm que declarar um motivo de entrada e explicar os seus objectivos para entrar e viver num dado país. Há vários motivos para entrar em Portugal, tendo-se registado os seguintes no Distrito de Évora: trabalho, reagrupamento familiar, estudo, reforma e outro. O motivo mais frequente é o reagrupamento familiar com 53%, a maioria destes são reagrupamentos entre os imigrantes do sexo feminino com o seu cônjuge, a seguir os filhos com os pais e os pais que se reagrupam com os filhos que já estão a viver em Portugal. O motivo de trabalho representa 30,7%. A seguir surge o motivo de estudo, com 6,1% dos respondentes. O motivo de reformado é muito baixo sendo de 1,6%, estes imigrantes são da União Europeia. Os outros motivos são 8,6%. Ver Tabela 7 e Figura 12.

Tabela 7. Distribuição de imigrantes por motivo de entrada

Motivo de Entrada		
	Frequência	Percentagem
Reagrupamento familiar	226	53
Trabalho	105	30,7
Outro	10	8,6
Estudo	5	6,1
Reforma	1	1,6
Total	313	100
NA	41	
Total	354	

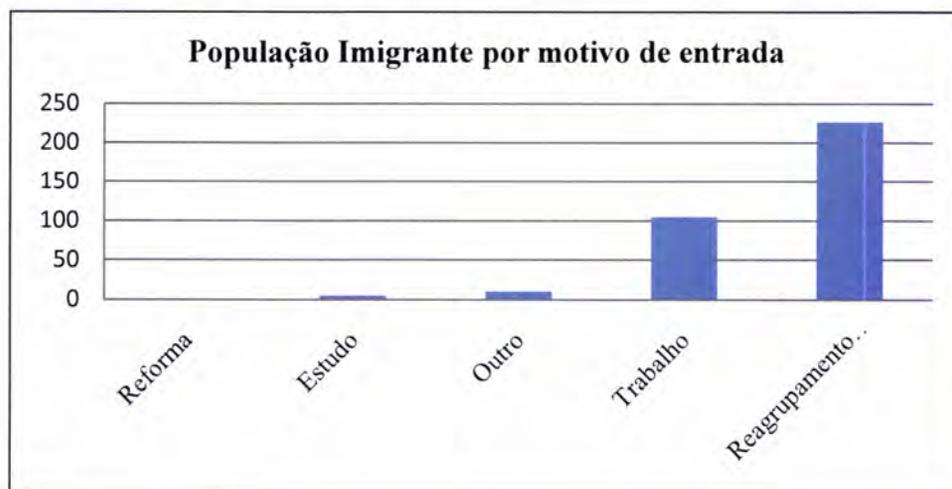


Figura 12. Diagrama de barras por motivo de entrada

❖ Profissão dos imigrantes

Recorrendo aos dados do questionário que foi aplicado pelo SEF é possível saber também a ocupação ou profissão dos imigrantes, verificando-se que os que têm emprego são 35,3%, na maioria trabalham nas obras, e como empregadas em lares dos idosos. A seguir aparecem os domésticos que são 29,3%; destes, a maioria são mulheres que casaram com portugueses. Estudantes são 18,9%, e podemos dizer que a ocupação de estudante é metade dos que têm emprego. Os desempregados representam-se 8,2%, e os reformados representam 3,5%. À procura do primeiro trabalho e outras profissões representam 1,9% e 2,8%, respectivamente. (Tabela 8 e Figura 13).

O conjunto de activos e de trabalhadores-estudantes representa 118 imigrantes, ou seja, 37,2%. O reduzido número de casos não permite fazer grandes generalizações.

Tabela 8. Distribuição de imigrantes por profissão

Profissão dos imigrantes		
	Frequência	Percentagem
Empregado	112	35,3
Doméstico	93	29,3
Estudante	60	18,9
Desemprego	26	8,2
Reformado	11	3,5
Outro	9	2,8
À procura do 1º trabalho	6	1,9
Total	317	100
NA	37	
Total	354	

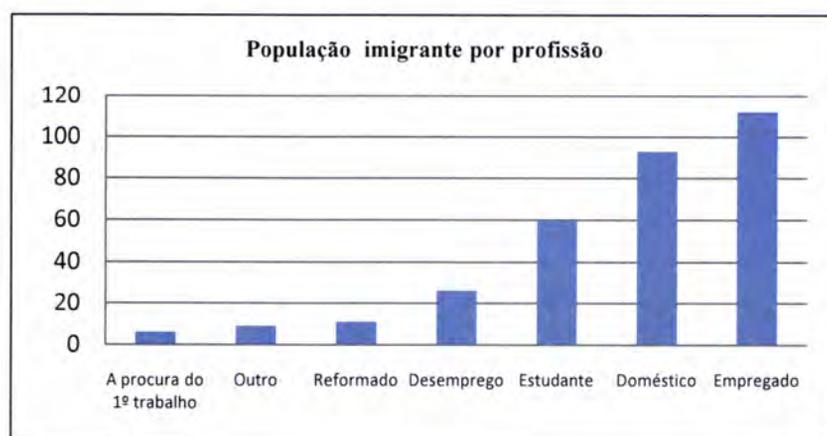


Figura 13. Diagrama de barra por Profissão

❖ Tipo de celebração do casamento

Na Tabela 9 e Figura 14 pode ver-se a distribuição dos diferentes tipos de celebração. Podemos distinguir sete tipos de celebração do casamento que ocorreram nos processos de imigrantes comunitários no SEF em Évora A maioria dos casais celebraram um casamento de tipo Civil - sem convenção antenupcial, correspondendo a 45%; de seguida, surge o casamento Civil representando 21,1%. A União de Facto representa 16,5% no que respeita aos imigrantes que vivem juntos com o seu cônjuge há pelo menos três anos vivendo em conjunto ou com papel assinado a declararem essa união). Em seguida o casamento Civil - sob regime de comunhão parcial representa 7,8% e Civil - com convenção antenupcial, 5%. Finalmente, os que celebraram casamento tipo Civil - no regime adquiridos e Católico - sem convenção antenupcial representam 2,3%.

Tabela 9. Distribuição por tipo de celebração do casamento

Tipo de Celebração do casamento		
	Frequência	Percentagem
Civil - sem convenção antenupcial	98	45
Civil	46	21,1
União de facto	36	16,5
Civil - sob regime da comunhão parcial	17	7,8
Civil - com convenção antenupcial	11	5
Civil - no regime adquiridos	5	2,3
Católico - sem convenção antenupcial	5	2,3
Total	218	100
NA	136	
Total	354	

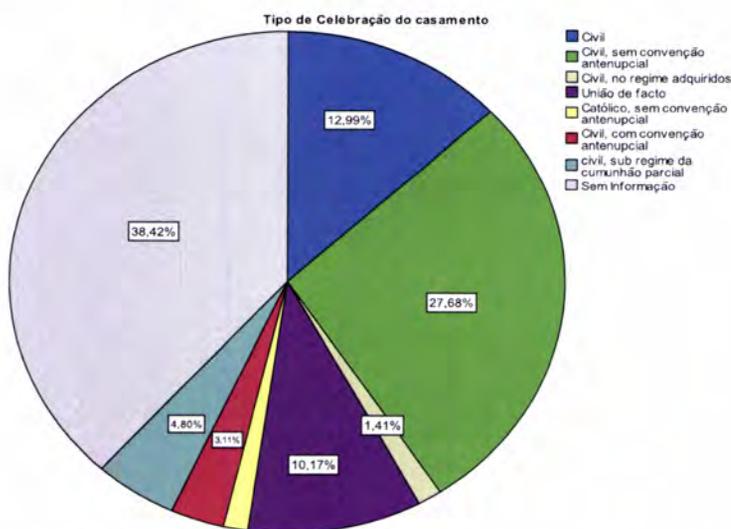


Figura 14. Diagrama circular por tipo de celebração do casamento

❖ União Europeia

A maioria dos imigrantes vem de países que não fazem parte da União Europeia, que representa 63,6%, isto é, aproximadamente 2/3 dos imigrantes que entram no Distrito de Évora são de fora da União Europeia (Tabela 10 e Figura 15).

Tabela 10. Distribuição por proveniência dentro ou fora da União Europeia

União Europeia		
	Frequência	Porcentagem
Não	225	63,6
Sim	129	36,4
Total	354	

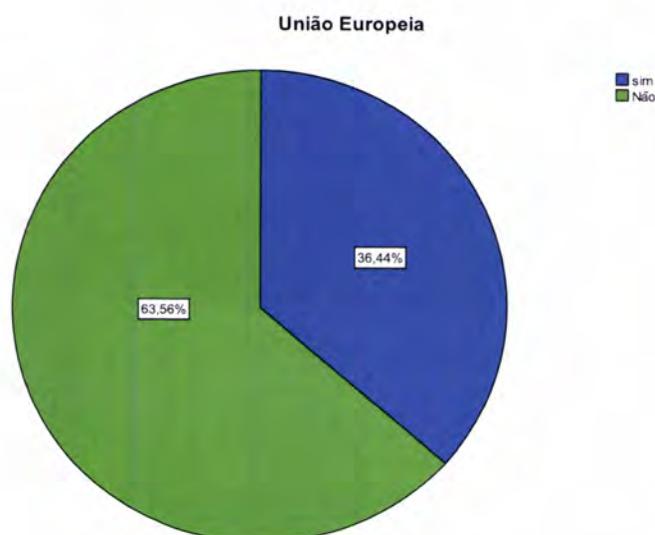


Figura 15. Diagrama de circular por proveniência dentro ou fora da União Europeia

❖ Naturalidade dos imigrantes

Observa-se que maioria dos imigrantes nascem nos países de origem, como é o caso do Brasil. A questão da distinção entre a nacionalidade e a naturalidade é importante pois mostra, entre outras dimensões, a diversidade dos percursos migratórios. A diferença verificada nos números entre elas relaciona-se com a aquisição de nacionalidade por outro meio que não o nascimento no país. Outra estratégia de aquisição da nacionalidade é o casamento, guerra, reagrupamento familiar. Mesmo assim os imigrantes na sua maioria têm nacionalidade igual ao país de origem (Tabela 11 e Figura 16).

Tabela 11. Distribuição por naturalidade dos imigrantes

Naturalidade dos imigrantes		
	Frequência	Porcentagem
Brasil	165	46,6
Roménia	28	7,9
Holanda	20	5,6
Alemanha	19	5,4
Espanha	17	4,8
Ucrânia	11	3,1
Itália	11	3,1
Bulgária	9	2,5
Cabo-Verde	7	2,0
Lituana	7	2,0
Moçambique	5	1,4
Angola	5	1,4
França	3	0,8
América	3	0,8
São Tome	3	0,8
Portugal	3	0,8
Belga	3	0,8
Marrocos	3	0,8
Hungria	3	0,8
Costa Marfim	2	0,6
Guiné-Bissau	2	0,6
Peru	2	0,6
Moldova	2	0,6
China	2	0,6
Venezuela	2	0,6
Rússia	2	0,6
Rep. Checa	2	0,6
Reino Unido	2	0,6
Bolívia	1	0,3
Colômbia	1	0,3
Nova Zelândia	1	0,3
Vietname	1	0,3
Filipina	1	0,3
Argentina	1	0,3
Suíça	1	0,3
Finlândia	1	0,3
Costa Rica	1	0,3
Estónia	1	0,3
Irlanda	1	0,3
Total	354	100,0

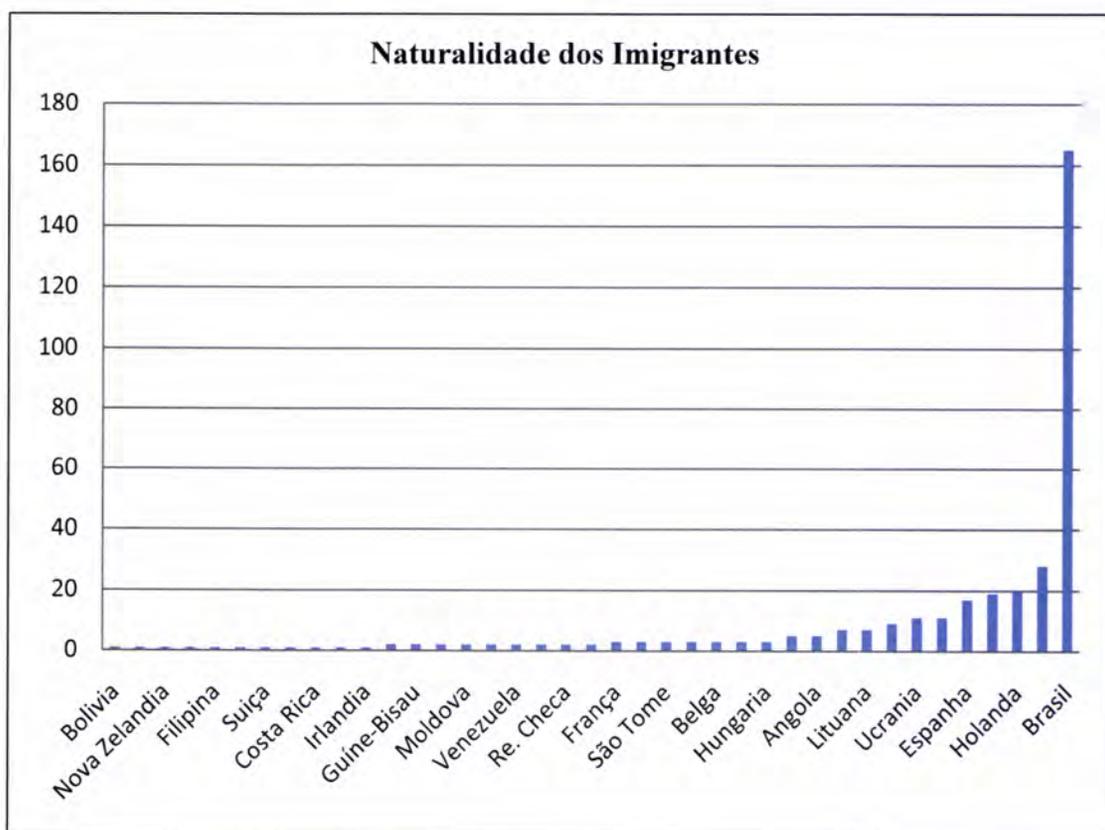


Figura 16. Diagrama de barra por naturalidade

❖ Naturalidade com codificação por classes

Dado o elevado número de países na dimensão naturalidade, resolvemos codificar a naturalidade dos imigrantes, tal como se fez para a nacionalidade. No caso brasileiro o país é igual à classe, nos outros casos criaram-se classes à semelhança do que se fez para a nacionalidade. A naturalidade menos expressiva é África que representa 6,1% e a mais significativa é Brasil que representa 53% (Tabela 12 e Figura 17).

Tabela 12. Distribuição dos imigrantes por Codificação da nacionalidade por classes

Codificação Naturalidade dos imigrantes		
	Frequência	Percentagem
Brasil	165	53,0
Europa	111	30,7
Restantes	54	8,6
África	24	6,1
Total	354	1,6

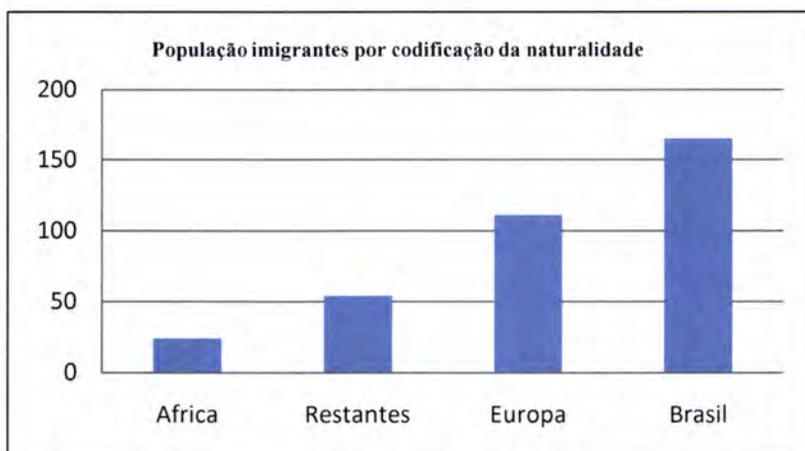


Figura 17. Diagrama de barra por codificação da naturalidade

❖ Naturalidade e Nacionalidade adquiridas

Cerca de 98% dos indivíduos têm nacionalidade idêntica à naturalidade. A seguir surgem imigrantes nascidos num país mas que adquiriram outra nacionalidade, apenas, 2%. Há 348 indivíduos que mantiveram a nacionalidade de origem e 6 indivíduos que adquiriram outra nacionalidade.

No que respeita à nacionalidade, verifica-se que três indivíduos nasceram em Portugal e adquiriram outra nacionalidade. Apenas um indivíduo foi adquirir a nacionalidade Portuguesa. Os outros dois têm a naturalidade de Marrocos e Moçambique também foram adquirir diferentes nacionalidades (Ver na tabela 13 e Figura 18).

Tabela 13. Distribuição de imigrantes por diferença entre naturalidade e nacionalidade dos imigrantes

Id	Naturalidade	Adquirir Nacionalidade
1	Portugal	Francesa
2	Portugal	Polaco
3	Portugal	Romena
4	Angola	Portuguesa
5	Marrocos	Holandês
6	Moçambique	Cabo-verdiana



Figura 18. Diagrama de barra para os imigrantes que mantêm nacionalidade e que adquiriram outra nacionalidade

❖ Naturalidade dos cônjuges

A maioria dos cônjuges nasceram em Portugal, representando 70 %; em seguida no Peru, 4,5%, Angola, 3,6%, e Holanda, 3,1%. As nacionalidades menos representativas são: Ucrânia, China, Romena, Costa Marfim e outros, apenas 0,4%. Finalmente os que representam 1,8% são Brasil, Moçambique, Itália, Argentina, Alemanha e Rússia, (Tabela 14 e Figura 19).

Tabela 14. Distribuição da nacionalidade dos cônjuges

Naturalidade do Cônjuge		
	Frequência	Porcentagem
Portugal	156	70,0
Peru	10	4,5
Angola	8	3,6
Holanda	7	3,1
Lituana	5	2,2
Brasil	4	1,8
Moçambique	4	1,8
Itália	4	1,8
Argentina	4	1,8
Alemanha	4	1,8
Rússia	4	1,8
América	2	0,9
Hungria	2	0,9
Cabo-Verde	1	0,4
Ucrânia	1	0,4
Costa Marfim	1	0,4
Guine	1	0,4
França	1	0,4
Roménia	1	0,4
China	1	0,4
Venezuela	1	0,4
Espanha	1	0,4
Total	223	
NA	131	
Total	354	

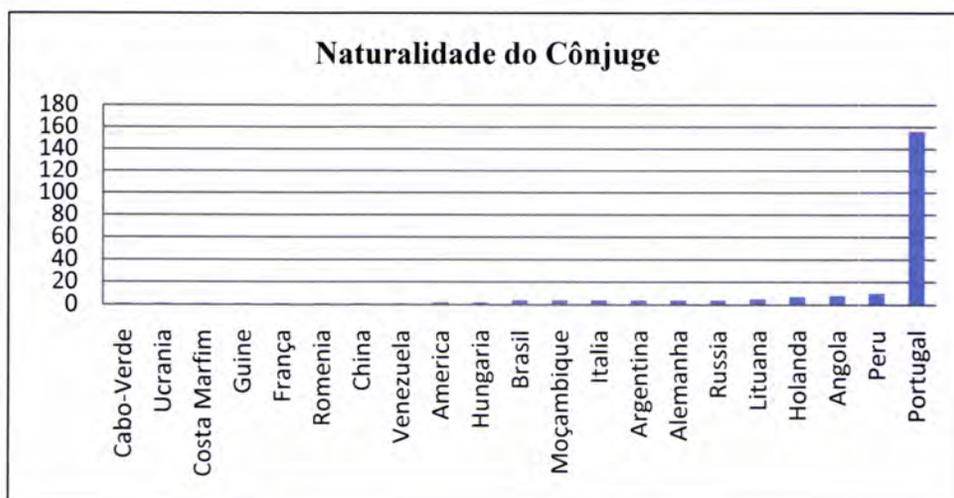


Figura 19. Diagrama de barras por nacionalidade dos cônjuges

❖ Nacionalidade dos cônjuges

Na Tabela 15 e Figura 20 pode observar-se a totalidade da nacionalidade dos cônjuges, correspondendo a 227 indivíduos. Registamos 15 nacionalidades e relativamente a 127 indivíduos não há informação nos anos de 2006 a 2009. Observa-se que a nacionalidade dos cônjuges é na sua maioria portuguesa com uma percentagem de 76,7% dos 227 indivíduos. A seguir surge as nacionalidades romena representando 5,3% e holandesa 4,4%. Finalmente, as outras nacionalidades apresentam um valor residual.

Tabela 15. Distribuição da nacionalidade dos cônjuges

Nacionalidade do Cônjuge		
	Frequência	Percentagem
Portuguesa	174	76,7
Romena	12	5,3
Holandesa	10	4,4
Alemã	9	4,0
Lituana	4	1,8
Búlgara	4	1,8
Brasileira	3	1,3
Italiana	3	1,3
Americana	2	0,9
Ucraniana	1	0,4
Moçambicana	1	0,4
Peruana	1	0,4
Angolana	1	0,4
Chinesa	1	0,4
Espanhola	1	0,4
Total	227	100,0
NA	127	
Total	354	

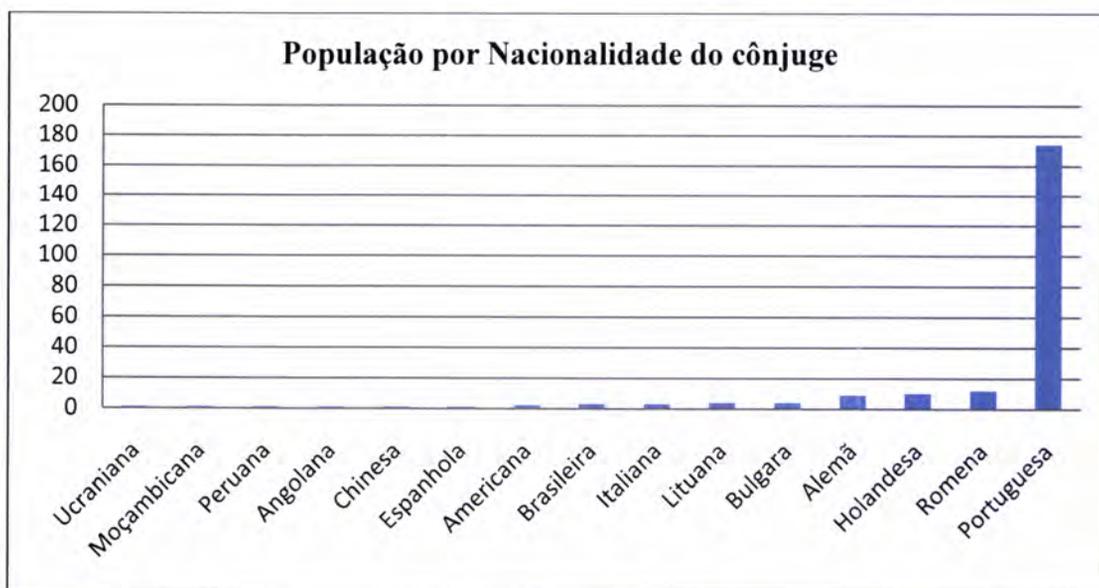


Figura 20. Diagrama de barra por nacionalidade do cônjuge

❖ Profissão dos cônjuges

Verifica-se que 89,3% dos cônjuges têm emprego, outros representam 5,1%. A seguir aparecem os reformados que são 2,8%. A ocupação como doméstico e estudante representam 1,1%, e os desempregados apenas 0,6%. Podemos concluir-se a maioria dos cônjuges têm emprego (ver Tabela 16 e Figura 21).

Tabela 16. Distribuição por profissão dos cônjuges

Profissão do Cônjuge		
	Frequência	Porcentagem
Empregado	158	89,3
Outro	9	5,1
Reformado	5	2,8
Doméstico	2	1,1
Estudante	2	1,1
Desemprego	1	0,6
Total	177	100,0
NA	177	
Total	354	

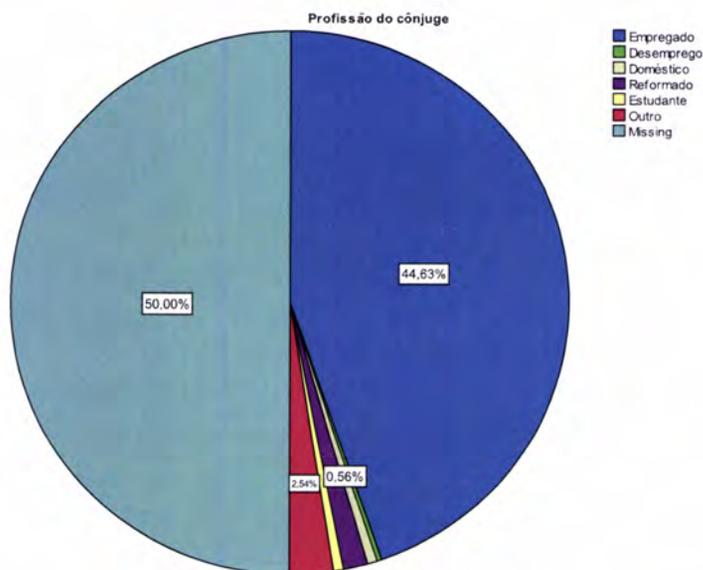


Figura 21. Diagrama circular por tipo de profissão dos cônjuges

❖ Idade do cônjuge

Na Tabela 17 observa-se que os cônjuges têm uma idade média aproximadamente a 43 anos, superior à idade de média dos imigrantes (36 anos). Metade dos cônjuges dos imigrantes têm 42 anos ou menos. Verifica-se na caixa de bigodes que existem três outliers (que são os indivíduos 195, 241 e 306 que tem respectivamente 77, 76 e 97 anos). Ver na Tabela 17 e na Figura 22 e 23.

Tabela 17. Estatística para idade dos cônjuges

Estatística		
Idade dos Cônjuges		
N	221	
Média	42,70	
Mediana	42,00	
Moda	48,00	
Desvio Padrão	12,06	
Variância	145,37	
Mínimo	20	
Máximo	97	
Percentis	25	33,00
	50	42,00
	75	49,50

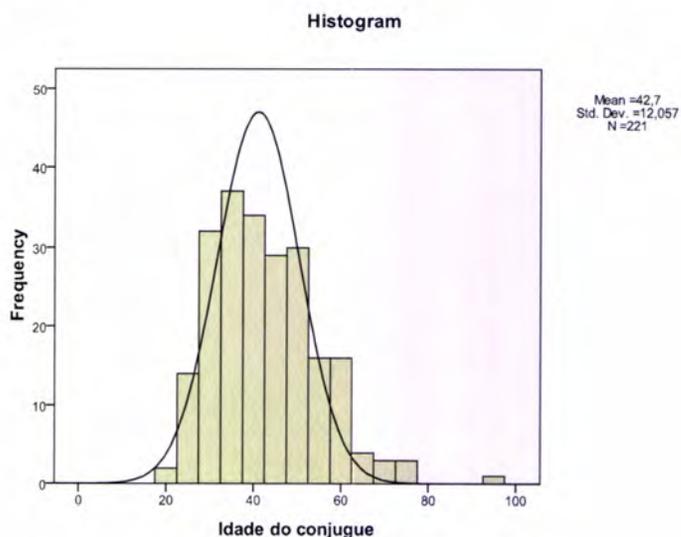


Figura 22. O histograma para idade do cônjuge

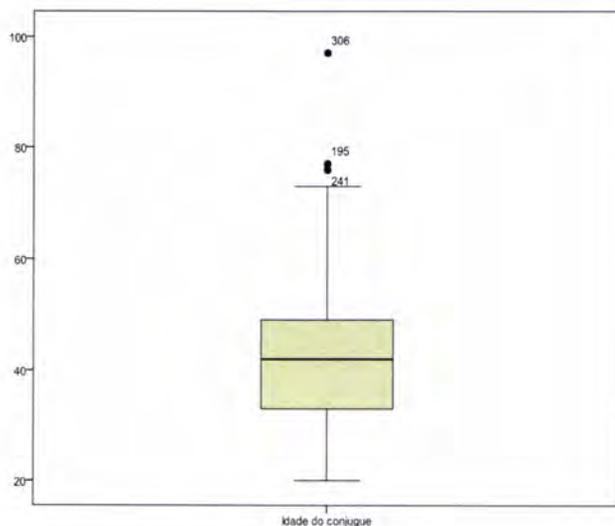


Figura 23. Caixa de bigodes para idade do cônjuge

❖ Número de filhos do casamento

Como a quantidade dos filhos se situa apenas entre 0 a 4 então podemos utilizar a tabela de frequências. Na maioria dos países europeus, o número de filhos tido como ideal tem vindo a tornar-se cada vez menor, e varia de um a dois ou, no máximo, três filhos por casal. Observa-se na Tabela 18 que, em média, o número de filhos dos casamentos é 1, sendo que este representa 41,1%. Cerca de 28,9 % representam os que não têm filhos e repare-se também que apenas menos de 25% dos imigrantes tiveram mais de 2 filhos. Os que tiveram 4 filhos é menos representativa, só 1,1%.

Tabela 18. Frequência para números de filhos do casamento

Números de filhos do casamento		
Nº Filhos	Frequência	Percentagem
1	37	41,1
0	26	28,9
2	22	24,4
3	4	4,4
4	1	1,1
Total	90	100
NA	264	
Total	354	

Em seguida representa-se a diagrama de barras para o número de filhos do casamento. A figura 24 também mostra que a maioria tem apenas um filho. A barra do diagrama para os que não têm filhos é ainda maior do que a dos que têm dois filhos.

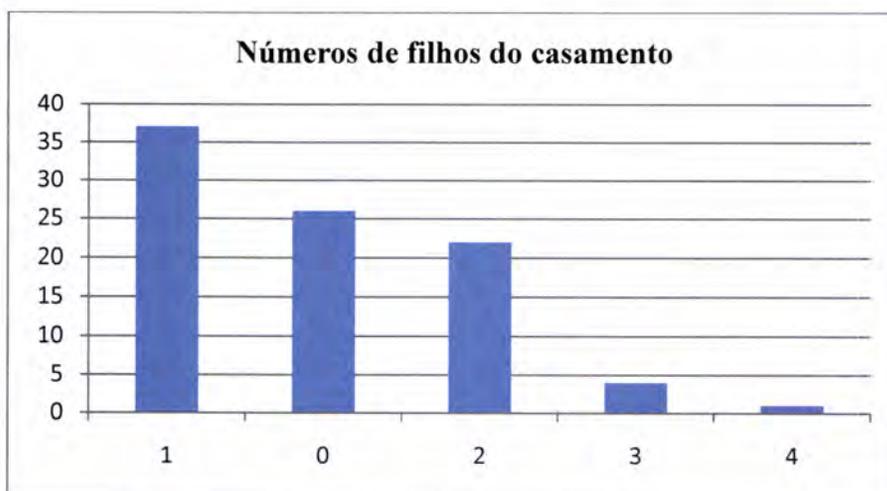


Figura 24. Diagrama de barra para número de filhos do casamento

❖ Números de filhos de outro casamento

O número de filhos de outro casamento não é uma variável contínua, como estas variáveis tomam poucos valores então aplica-se a tabela de frequências. Os filhos de outro casamento são os que foram trazidos pela mãe quando causou com o primeiro marido. Quando termina o primeiro casamento então os filhos ficam com a mãe, pelo que esta quando casar novamente fica com os filhos na sua responsabilidade e do novo marido. A maioria, são os que não têm filhos de outro casamento, representa 47,3% e cerca de 38,2% os que têm um filho. Finalmente os que têm dois filhos e três filhos representam 12,7% e 1,8% (Tabela 19 e Figura 25).

Tabela 19. Frequência para número de filhos de outro casamento

Número de filhos de Outro Casamento		
Nº Filhos	Frequência	Percentagem
0	26	47,3
1	21	38,2
2	7	12,7
3	1	1,8
Total	55	100
NA	299	
Total	354	

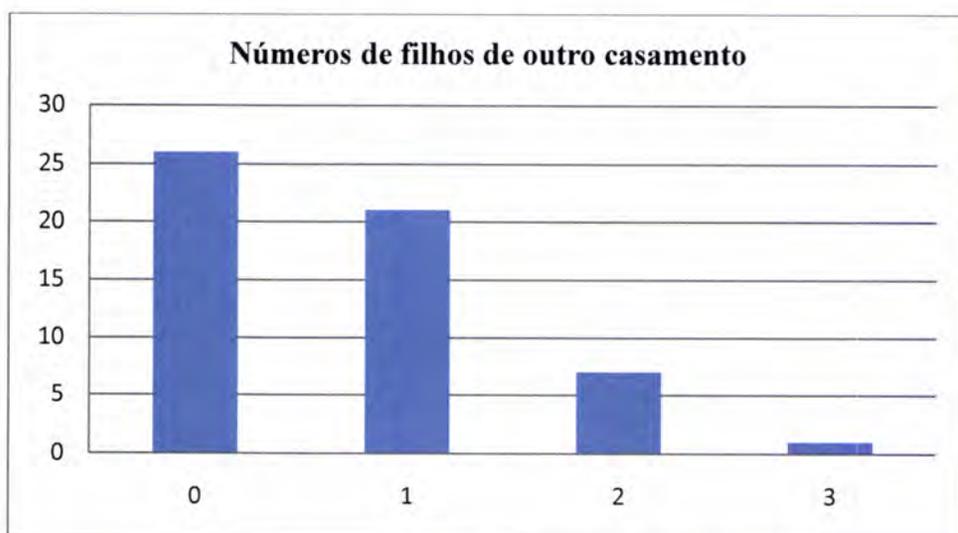


Figura 25. Diagrama de barras por números de filhos de outro casamento

3.3. Associação de Variáveis

A análise estatística nesta secção permite-nos analisar simultaneamente duas ou mais variáveis, averiguando se existem relações significativas entre algumas variáveis.

As variáveis que consideramos são de natureza nominal e os valores referem-se ao número ou frequência de casos que se situam em cada categoria, estabelecendo tabelas de contingência ou cruzadas. Nas tabelas de contingência podem cruzar-se variáveis pertencentes ao mesmo fenómeno, de modo a perceber as características mais importantes que o compõem, ou cruzar-se variáveis pertencentes a fenómenos diferentes, de forma a verificar a existência de relações de dependência.

3.3.1. Relação Entre Sexo e País de Origem

Com base no Quadro 1 pretende-se analisar a relação entre as classes da União Europeia e o sexo dos imigrantes. Entre os imigrantes fora da União Europeia o sexo feminino apresenta um número maior (180) do que o sexo masculino (45) (Ver no Quadro 1). Ao contrário entre os imigrantes da União Europeia são os homens que predominam.

Quadro 1. Relação entre o sexo e países de origem

Sexo * União Europeia Crosstabulation

			União Europeia		Total
			sim	Não	
Sexo	Feminino	Count	62	180	242
		Expected Count	88,2	153,8	242,0
		Std. Residual	-2,8	2,1	
	Masculino	Count	67	45	112
		Expected Count	40,8	71,2	112,0
		Std. Residual	4,1	-3,1	
Total	Count	129	225	354	
	Expected Count	129,0	225,0	354,0	

De acordo com o Quadro 1 podemos dizer que os resíduos ajustados na forma estandardizada informam sobre as células que mais se afastam da hipótese de independência entre as variáveis. Vêm mais da União Europeia imigrantes do sexo masculino do que o esperado. Por outro lado, os que não fazem parte da União Europeia vêm mais do sexo feminino, portanto, esses resíduos serão negativos porque o valor esperado maior do que valor observado.

Quadro 2. Estatística teste do qui-quadrado

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	38,668 ^a	1	<0,0001		
Continuity Correction ^b	37,206	1	,000		
Likelihood Ratio	38,053	1	,000		
Fisher's Exact Test				,000	,000
Linear-by-Linear Association	38,559	1	,000		
N of Valid Cases	354				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 40,81.

b. Computed only for a 2x2 table

Pelo teste do qui-quadrado obtêm-se um valor $p = 0.0001 < \alpha = 0.1$, pelo que existe evidência estatística para concluir que existe associação entre sexo e origem, sendo que neste caso há muito mais imigrantes do sexo feminino oriundos de fora da União Europeia do que haveria no caso de independência. Existe, pois, evidência estatística de que o género dos imigrantes não é independente dos países de origem.

Quadro 3. O Risco Estimado

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Sexo (Feminino / Masculino)	,231	,144	,372
For cohort País União Europeia = sim	,428	,329	,557
For cohort País União Europeia = Não	1,851	1,459	2,348
N of Valid Cases	354		

De acordo com o Quadro 3 (O Risco Estimado), os indivíduos do sexo masculino têm uma possibilidade 4 vezes superior de pertencerem a um país da União Europeia.

3.3.2. Relação Entre o Ano de Entrada em Évora e a Nacionalidade

Para estudar a relação entre o ano de entrada em Évora e a nacionalidade, pode-se analisar o quadro seguinte:

Quadro 4. Ano de entrada em Évora e nacionalidade

Ano entrada * Codificação Nacionalidade Crosstabulation

		Codificação Nacionalidade				Total
		Brasil	África	União Europa	Restantes	
Ano entrada 2006	Count	21	2	42	14	79
	Expected Count	36,8	4,9	24,8	12,5	79,0
	Std. Residual	-2,6	-1,3	3,5	,4	
2007	Count	47	7	65	19	138
	Expected Count	64,3	8,6	43,3	21,8	138,0
	Std. Residual	-2,2	-,5	3,3	-,6	
2008	Count	57	7	3	17	84
	Expected Count	39,2	5,2	26,3	13,3	84,0
	Std. Residual	2,9	,8	-4,5	1,0	
2009	Count	40	6	1	6	53
	Expected Count	24,7	3,3	16,6	8,4	53,0
	Std. Residual	3,7	1,5	-3,8	-,8	
Total	Count	165	22	111	56	354
	Expected Count	165,0	22,0	111,0	56,0	354,0

Em relação ao ano de entrada, no ano 2006 entraram mais imigrantes que fazem parte da União Europeia, também foi assim em 2007. Mas, em 2008 e 2009 em relação à codificação da nacionalidade a maioria é brasileira (Quadro 4). O maior resíduo é (3,5) respeita aos Europeus que entraram em 2006, o que significa que dos 111 imigrantes Europeus se esperava 24,8 consideram entraram no ano 2006. Porém observam-se 42 imigrantes que entraram 2006 e 2007, valores bastantes parecidos e, por outro lado, também parecidos com os brasileiros que entraram no ano 2009; representa-se 3,1 indicam as células que se afastam da hipótese de independência.

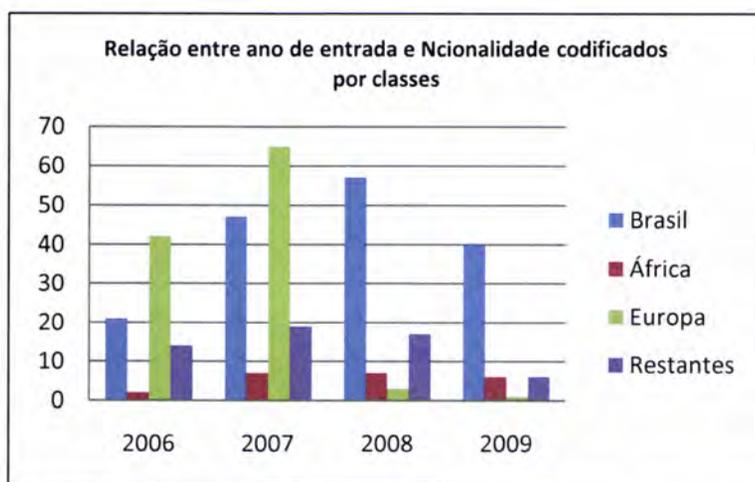


Figura 26. O diagrama de barras em relação ano de entrada e nacionalidade codificados por classes

Observa-se na Figura 26, a relação entre ano de entrada e nacionalidade codificados por classes, o número de brasileiros no ano 2006, 2007, 2008 cresceu bastante, mas no ano de 2009 diminuiu.

Quadro 5. Estatística teste para o qui-quadrado

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	94,431 ^a	9	<0,0001
Likelihood Ratio	114,016	9	,000
Linear-by-Linear Association	42,169	1	,000
N of Valid Cases	354		

a. 2 cells (12,5%) have expected count less than 5. The minimum expected count is 3,29.

Observamos que no Quadro 5 as combinações entre ano de entrada e os países codificados têm algumas células menores do que 5. Observa-se o valor $p = 0.000 < \alpha = 0.05$, então rejeita-se a hipótese pois é altamente significativa, ou seja, existe evidência estatística para afirmar que as nacionalidades codificadas por classes são dependentes do ano de entrada.

Para confirmar o decréscimo de imigrantes da União Europeia nos anos 2007, 2008 e 2009 fez-se um pedido de informação à Câmara Municipal de Évora relativamente ao número de certificados que foram emitidos nesses anos.

De acordo com a informação recebida da Câmara Municipal de Évora foram 156 os certificados para os imigrantes que fazem parte da União Europeia emitidos nos anos de 2007 a 2009. Não existem dados para o ano 2006 porque o município só começou a emitir os referidos certificados em 2007 (Tabela 20). Assim, podemos juntar os imigrantes da União Europeia que foram tratar do título da residência no SEF e os imigrantes que foram pedir os certificados na Câmara Municipal.

Observa-se na Tabela 20 os números resultantes da junção dos imigrantes da União Europeia que foram pedir os certificados na Câmara Municipal de Évora e os que foram tratar o título da residência no SEF. Para os imigrantes que fazem parte da União Europeia no ano 2006 o valor mantém-se porque a Câmara Municipal ainda não emitia certificados neste ano, os valores para 2007, 2008 e 2009 aumentam.

Tabela 20. Os imigrantes que foram certificados na Câmara Municipal

Os imigrantes da União Europeia que foram emitir certificados na Câmara Municipal		
	Números certificados	
Ano entrada	2006	-
	2007	19
	2008	92
	2009	45
Total		156

Fonte: Câmara Municipal de Évora

Depois de somados os números referidos anteriormente, verifica-se que a maioria dos imigrantes que entraram para Évora são imigrantes que fazem parte da União Europeia. De seguida, aparecem os imigrantes brasileiros, os imigrantes do grupo “restantes países” e, finalmente, os imigrantes Africanos que representam a minoria dos imigrantes que entraram em Évora no período de 2006 a 2009.

Tabela 21. Relação entre o ano de entrada em Évora e a nacionalidade codificada (Juntando a União Europeia pelo SEF e Município)

Ano entrada * Codificação Nacionalidade Crosstabulation						
		Codificação Nacionalidade				Total
		Brasileiras	Africanas	União Europa	Restantes Nacionalidade	
Ano entrada	2006	21	2	42	14	79
	2007	47	7	84	19	157
	2008	57	7	95	17	176
	2009	40	6	46	6	98
Total		165	22	267	56	510

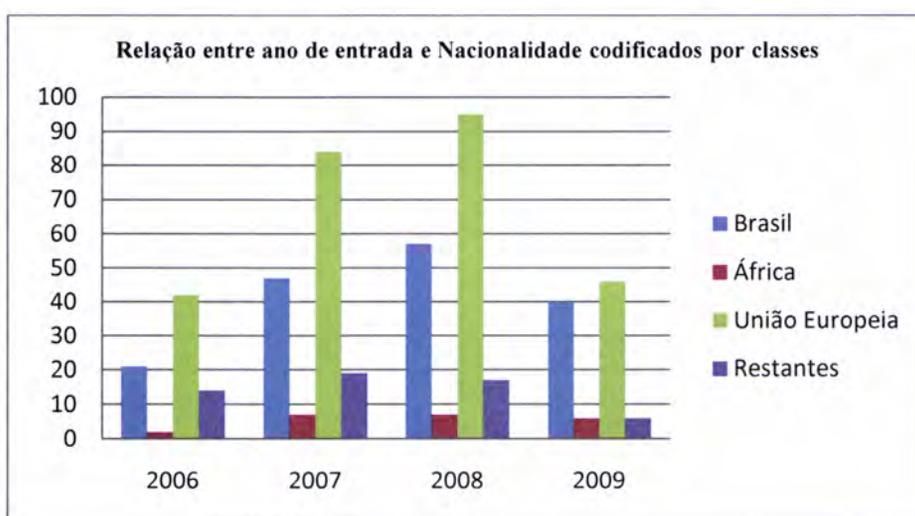


Figura 27. Diagrama de barras em relação ao ano de entrada e nacionalidade codificada por classes (Juntando a União Europeia pelo SEF e Município)

Verifica-se na Figura 27, a relação entre ano de entrada e nacionalidade, codificada por classes. O número de imigrantes da União Europeia nos anos 2006, 2007 e 2008 cresceu bastante, mas no ano 2009 diminuiu. Para o continente Africano cresceu entre 2006 e 2007, e no ano 2008 e 2009 mantém-se. Os imigrantes de nacionalidade brasileira que entraram em Évora cresceu entre 2006 e 2008, mas no ano 2009 decresceu. Os imigrantes dos países restantes cresceram bastante nos anos 2006 e 2007, nos anos 2007 e 2008 diminuíram ligeiramente, tendência que se acentuou entre 2008 a 2009.

Quadro 6. Relação entre país de origem e ano de entrada

Origem * Ano de Entrada Crosstabulation

			Ano de Entrada				Total
			2006	2007	2008	2009	
Origem	Brasileira	Count	21	47	57	40	165
		Expected Count	25,6	50,8	56,9	31,7	165,0
		Std. Residual	-,9	-,5	,0	1,5	
	Africanas	Count	2	7	7	6	22
		Expected Count	3,4	6,8	7,6	4,2	22,0
		Std. Residual	-,8	,1	-,2	,9	
	União Europa	Count	42	84	95	46	267
		Expected Count	41,4	82,2	92,1	51,3	267,0
		Std. Residual	,1	,2	,3	-,7	
3,00	Count	14	19	17	6	56	
	Expected Count	8,7	17,2	19,3	10,8	56,0	
	Std. Residual	1,8	,4	-,5	-1,5		
Total	Count	79	157	176	98	510	
	Expected Count	79,0	157,0	176,0	98,0	510,0	

Quadro 7. Estatística teste para qui-quadrado

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	11,167 ^a	9	,264
Likelihood Ratio	10,990	9	,276
Linear-by-Linear Association	7,679	1	,006
N of Valid Cases	510		

a. 2 cells (12,5%) have expected count less than 5. The minimum expected count is 3,41.

De acordo com o Quadro 6 mostra-se que os valores das células do observado e esperado são aproximadamente iguais, então podemos dizer que na maior parte os números dos imigrantes observados são parecidos com os números dos imigrantes esperados. Observa-se nos valores das células e do Quadro 7 com o valor de $p = 0.264 > \alpha = 0.1$, significa que não existe uma relação dependência entre o ano de entrada e o país de origem.

3.3.3. Relação Entre Designação da Morada e Países União Europeia

Estudámos a relação entre a origem dos imigrantes e a sua designação da morada (rural ou urbana).

Quadro 8. Tabela de contingência entre designação da morada e países da União Europeia

Designação da Morada * País União Europeia Crosstabulation

Count

		País União Europeia		Total
		sim	Não	
Designação da Morada	Rural	54	103	157
	Urbano	66	119	185
Total		120	222	342

Observa-se que os imigrantes não europeus tem um maior número de designação da morada em meio urbano (119) do que os que vivem no meio rural (103) (Ver Quadro 8). A relação entre o país de origem e a designação da morada caracteriza-se ainda pelo facto dos imigrantes que fazem parte da União Europeia também residirem mais em meio urbano.

Quadro 9. Estatística Teste do Qui-quadrado

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,061 ^a	1	,805		
Continuity Correction ^b	,018	1	,894		
Likelihood Ratio	,061	1	,805		
Fisher's Exact Test				,821	,447
Linear-by-Linear Association	,061	1	,805		
N of Valid Cases	342				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 55,09.

b. Computed only for a 2x2 table

Observa-se no Quadro 9 a estatística teste do qui-quadrado para ver se associação das variáveis dos países de origem e designação da morada é significativa ou não. Nota-se o valor $p = 0.805 > \alpha = 0.05$, então não se rejeita a hipótese de independência porque a relação não é significativa. Concluímos que não existe evidência estatística para afirmar que a designação da morada depende do país de origem.

3.3.4. Sexo e Designação da Morada

A importância desta relação para ver se os imigrantes que foram casar com os portugueses vivem mais no meio rural ou no urbano, em função do seu género. A tabela de contingência obtida foi a seguinte:

Quadro 10. Tabela de contingência entre sexo e designação da morada

Sexo * Designação da Morada Crosstabulation					
			Designação da Morada		Total
			Rural	Urbano	
Sexo	Feminino	Count	110	126	236
	Masculino	Count	47	59	106
Total		Count	157	185	342

Em seguida apresenta-se o teste qui-quadrado para ver a independência entre as variáveis.

Quadro 11. Estatística teste do qui-quadrado

Chi-Square Tests					
	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,152 ^a	1	,697		
Continuity Correction ^b	,074	1	,785		
Likelihood Ratio	,152	1	,697		
Fisher's Exact Test				,726	,393
Linear-by-Linear Association	,151	1	,697		
N of Valid Cases	342				

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 48,66.

b. Computed only for a 2x2 table

Em relação ao sexo e designação da morada observa-se um maior número de imigrantes do sexo feminino que residem no meio urbano (Quadro 10). Assim sendo, os que vivem no rural, também do sexo feminino, os valores das células não existe grande diferença.

O resultado do Quadro 11 corresponde ao teste qui-quadrado, observando-se um valor $p = 0.697 > \alpha = 0.05$, então não se rejeita a hipótese de independência porque a relação não é significativa. Concluímos que a designação da morada é independente do sexo.

3.3.5. Relação Entre Nacionalidade e Sexo

Observa-se no Quadro 12 a relação entre o sexo e a nacionalidade codificada por classes. Existe um grande número de imigrantes de nacionalidade Brasileira do sexo feminino, o que se verifica também com os oriundos de África. Em seguida os que tem nacionalidade da Europa do sexo masculino.

Quadro 12. Tabela de contingência da codificação da nacionalidade e o sexo

Codificação Nacionalidade * Sexo Crosstabulation

			Sexo		Total
			Feminino	Masculino	
Codificação Nacionalidade	Brasil	Count	134	31	165
		Expected Count	112,8	52,2	165,0
		Std. Residual	2,0	-2,9	
	África	Count	20	2	22
		Expected Count	15,0	7,0	22,0
		Std. Residual	1,3	-1,9	
	Europa	Count	55	56	111
		Expected Count	75,9	35,1	111,0
		Std. Residual	-2,4	3,5	
	Restantes	Count	33	23	56
		Expected Count	38,3	17,7	56,0
		Std. Residual	-,9	1,3	
Total	Count	242	112	354	
	Expected Count	242,0	112,0	354,0	

Quadro 13. Estatística teste do qui-quadrado

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	38,235 ^a	3	<0,001
Likelihood Ratio	39,328	3	,000
Linear-by-Linear Association	26,056	1	,000
N of Valid Cases	354		

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 6,96.

No Quadro 12 mostra-se que vieram mais brasileiras do que aquilo que esperava. Ao contrário, esperava-se que houvesse mais pessoas brasileiras do sexo masculino, razão pelo qual o resíduo é negativo. Em seguida os que vêm da Europa são mais do sexo masculino do

que aquilo que esperado; nota-se os resíduos estandardizados que são positivos. Por outro lado, os imigrantes de Europa do sexo feminino vêm muito menos do que aquilo que esperado.

Observa-se no Quadro 13 o teste qui-quadrado com valor $p = 0.001 < \alpha = 0.1$, pelo que podemos concluir que existe evidência estatística para afirmar que a nacionalidade está associada ao género. Portanto, há influência das diferenças de nacionalidade sobre o sexo.

3.3.6. Relação Entre o Estado Civil e o Sexo

A análise da relação entre o estado civil e sexo é feita pelo quadro seguinte:

Quadro 14. Tabela contingência do sexo e estado civil

Sexo * Estado civil Crosstabulation

Count

		Estado civil					Total
		Solteiro	Casado	Divorciada	Viúva	Separada	
Sexo	Feminino	52	173	6	4	1	236
	Masculino	53	53	4	1	0	111
Total		105	226	10	5	1	347

No Quadro 14 apresenta-se a Tabela de Contingência obtida entre as variáveis sexo e estado civil. Havendo muitas células com valor observado menor do que 5 juntámos categorias, divorciada, viúva, separada com uma categoria “outros” obtendo o Quadro 15.

Observa-se que os casados têm um maior número do sexo feminino (173) do que do sexo masculino (53). Os que têm menor número são dos outros estados civis do sexo masculino (5), ou seja os que têm outro estado civil os imigrantes que vêm parecidos com aqueles que esperava. Comparamos o valor observado e esperado e existe uma grande diferença dos solteiros do sexo masculino do que aquilo que esperava. Nota-se os solteiros do sexo feminino e dos casados do sexo masculino esperavam-se mais pessoas, o que justifica o resíduo estandardizado ser negativo. Com base nas diferenças do valor observado e esperado pode-se dizer que se rejeita a hipótese nula (Quadro 15).

Quadro 15. Tabela contingência do sexo e codificação do estado civil

Codificação Ec * Sexo Crosstabulation

			Sexo		Total
			Feminino	Masculino	
Codificação Ec	Solteiro	Count	52	53	105
		Expected Count	71,4	33,6	105,0
		Std. Residual	-2,3	3,3	
	Casado	Count	173	53	226
		Expected Count	153,7	72,3	226,0
		Std. Residual	1,6	-2,3	
	Outros	Count	11	5	16
		Expected Count	10,9	5,1	16,0
		Std. Residual	,0	,0	
Total	Count	236	111	347	
	Expected Count	236,0	111,0	347,0	

Em seguida apresenta-se o quadro o teste qui-quadrado para tomar a decisão, será que existe uma relação de dependência entre as variáveis estado civil e sexo.

Quadro 16. Estatística teste do qui-quadrado

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	24,071 ^a	2	<0,001
Likelihood Ratio	23,369	2	,000
Linear-by-Linear Association	17,807	1	,000
N of Valid Cases	347		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 5,12.

000

No Quadro 16 mostra-se o valor $p = 0.0001 < \alpha = 0.1$, pelo que é altamente significativa ou seja a associação entre o sexo e o estado civil. Pelo que existe evidência estatística para concluir que existe associação entre sexo e estado civil, sendo que neste caso há muito mais imigrantes do sexo feminino que são casados. Assim existe evidência estatística de que o estado civil dos imigrantes depende do seu género.

3.3.7. Relação Entre o Nível de Escolaridade e o Sexo

Para estudar a relação entre o nível escolaridade e o sexo obtida através do seguinte Quadro da tabela contingência composto por valor esperado e observado.

Quadro 17. Tabela de contingência do sexo e do nível de escolaridade

Nível da escolaridade * Sexo Crosstabulation

		Sexo		Total
		Feminino	Masculino	
Nível da escolaridade Nenhum	Count	5	1	6
	Expected Count	4,2	1,8	6,0
	Std. Residual	,4	-,6	
Básico	Count	60	16	76
	Expected Count	52,6	23,4	76,0
	Std. Residual	1,0	-1,5	
Secundário	Count	103	44	147
	Expected Count	101,8	45,2	147,0
	Std. Residual	,1	-,2	
Técnico profissional	Count	10	8	18
	Expected Count	12,5	5,5	18,0
	Std. Residual	-,7	1,0	
Superior	Count	45	30	75
	Expected Count	51,9	23,1	75,0
	Std. Residual	-1,0	1,4	
Total	Count	223	99	322
	Expected Count	223,0	99,0	322,0

Quadro 18. Estatística teste do qui-quadrado

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	8,561 ^a	4	,073
Likelihood Ratio	8,638	4	,071
Linear-by-Linear Association	7,589	1	,006
N of Valid Cases	322		

a. 2 cells (20,0%) have expected count less than 5. The minimum expected count is 1,84.

No Quadro 17 apresenta-se a Tabela de contingência da relação entre o nível de escolaridade e o sexo, os que possuem o secundário têm um valor mais alto do sexo feminino (103) do que do sexo masculino (44), e a seguir ao nível básico tem 60 indivíduos do sexo feminino (60) do que do sexo masculino (16). Os que não têm nível escolaridade ou nenhum têm menores números quando comparamos com outros níveis.

O valor do $p = 0.078 < \alpha = 0.1$, assim rejeita-se a hipótese, o que significa que essas duas variáveis são independentes, ou seja existe evidência estatística para afirmar que o nível de escolaridade é dependente do sexo dos imigrantes (Ver Quadro 18).

3.3.8. Relação Entre Sexo e Motivo de Entrada

Para ver esta relação acontece da mesma forma como a análise anterior, que tem o valor observado das células menor do que 5 (Tabela de contingência apresentada no Quadro 19), por isso precisamos de juntar novamente a categoria “Outro” resulta de juntar as categorias iniciais Reforma e Outro. O resultado de juntar os valores da células menores do que 5 mostra-se na Tabela de Contingência do Quadro 20.

Quadro 19. Tabela contingência do sexo e do motivo de entrada

Sexo * Motivo de entrada Crosstabulation

Count

	Motivo de entrada					Total
	Trabalho	Reagrupamento familiar	Estudo	Reforma	Outro	
Sexo Feminino	52	127	14	3	20	216
Masculino	44	39	5	2	7	97
Total	96	166	19	5	27	313

De acordo com Quadro 20 podemos comparar o valor observado e esperado, os que do sexo masculino vêm mais com motivo de trabalho do que aquilo que esperava (nota-se o valor de resíduo estandardizado é 2,6). Os imigrantes do sexo feminino aparecem mais com motivo de reagrupamento familiar quando comparado com os esperados. Ao contrário, em relação ao sexo masculino esperavam-se mais do que os que vieram.

Vimos que a relação entre o sexo e o motivo de entrada, os que têm motivos de reagrupamento familiar têm um número superior do sexo feminino (127) do que do sexo masculino (39); podemos dizer que o sexo feminino parece ter maior possibilidade ou maior número aos todos dos motivos de entrada quando comparado com os motivos de entrada do sexo masculino (Ver no Quadro 22).

Quadro 20. Tabela contingência do sexo e codificação do motivo de entrada

codificação Moterna * Sexo Crosstabulation

		Sexo		Total
		Feminino	Masculino	
codificação Moterna Trabalho	Count	52	44	96
	Expected Count	66,2	29,8	96,0
	Std. Residual	-1,8	2,6	
Reagrupamento Familiar	Count	127	39	166
	Expected Count	114,6	51,4	166,0
	Std. Residual	1,2	-1,7	
Estudo	Count	14	5	19
	Expected Count	13,1	5,9	19,0
	Std. Residual	,2	-,4	
Outro	Count	23	9	32
	Expected Count	22,1	9,9	32,0
	Std. Residual	,2	-,3	
Total	Count	216	97	313
	Expected Count	216,0	97,0	313,0

No Quadro 21 apresenta-se o output do Teste Qui-quadrado tendo-se obtido um valor $p = 0.02 < \alpha 0.1$, conseqüentemente, podemos concluir que existe evidência estatística para afirmar que o motivo de entrada depende do gênero.

Quadro 21. Estatística teste para qui-quadrado

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	14,568 ^a	3	,002
Likelihood Ratio	14,169	3	,003
Linear-by-Linear Association	5,537	1	,019
N of Valid Cases	313		

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 5,89.

3.3.9. Relação Entre Países de Origem e Motivo de Entrada

Quadro 22. Tabela de contingência do país da União Europeia e do motivo da entrada

País União Europeia * Motivo de entrada Crosstabulation

Count

		Motivo de entrada					Total
		Trabalho	Reagrupamento familiar	Estudo	Reforma	Outro	
País União Europeia	sim	48	29	11	5	19	112
	Não	48	137	8	0	8	201
Total		96	166	19	5	27	313

De acordo com Quadro 22, existem valores das células menores do que 5; por isso, precisamos de juntar os valores das células menores do que 5. Relativamente ao Quadro 23, o resultado de juntar os valores observados apresenta-se no Quadro 23.

Observa-se a relação entre os países da União Europeia e motivo de entrada, continuam os que têm motivos de reagrupamento familiar têm um número superior dos não Europeus (137) do que dos Europeus (29); em relação ao motivo de trabalho para Europeus e não Europeus mantém-se mesmo valor. O motivo de estudo tem número maior dos que fazem parte da União Europeia (11) do que eles que não fazem parte (8), também para os outros motivos (Ver Quadro 24).

Por outro lado, os valores de comparação entre o valor esperado e o observado existe grande diferença. Assim, podemos afirmar que a relação de dependência entre as variáveis observadas na amostra. No Quadro 24 mostra-se o output do Teste Qui-quadrado no qual se obteve um valor do $p = 0.000$ pelo que podemos concluir que existe evidência estatística para afirmar que o motivo de entrada depende do ser oriundo ou não de um país da União Europeia.

Quadro 23. Tabela contingência do país União Europeia e da codificação do motivo de entrada

codificação Moterna * País União Europeia Crosstabulation

			País União Europeia		Total
			sim	Não	
codificação Moterna	Trabalho	Count	48	48	96
		Expected Count	34,4	61,6	96,0
		Std. Residual	2,3	-1,7	
	Reagrupamento Familiar	Count	29	137	166
		Expected Count	59,4	106,6	166,0
		Std. Residual	-3,9	2,9	
	Estudo	Count	11	8	19
		Expected Count	6,8	12,2	19,0
		Std. Residual	1,6	-1,2	
	Outro	Count	24	8	32
		Expected Count	11,5	20,5	32,0
		Std. Residual	3,7	-2,8	
Total	Count	112	201	313	
	Expected Count	112,0	201,0	313,0	

Quadro 24. Estatística teste do qui-quadrado

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	58,132 ^a	3	,000
Likelihood Ratio	59,511	3	,000
Linear-by-Linear Association	4,383	1	,036
N of Valid Cases	313		

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 6,80.

3.3.10. Relação Entre País de Origem e Profissão dos Imigrantes

A relação entre o país de origem e a profissão pode ser analisada através dos resultados em baixo. Observa-se no quadro 25, existem alguns valores das células menores do que 5. Precisamos de juntar os valores das células menores do que 5. As categorias que fomos juntar são “À procura do 1º trabalho, Desemprego e Reformado” na categoria “Outro”. O resultado do juntar dos valores observados apresenta-se no quadro seguinte.

Quadro 25. Tabela de contingência do País da União Europeia e da profissão dos imigrantes

País União Europeia * Profissão dos imigrantes Crosstabulation

Count

		Profissão dos imigrantes							Total
		Empregado	A procura do 1º trabalho	Desemprego	Doméstico	Reformado	Estudante	Outro	
País União Europeia	sim	53	2	4	10	8	21	5	103
	Não	59	4	22	83	3	39	4	214
Total		112	6	26	93	11	60	9	317

Quadro 26. Tabela de contingência do país da União Europeia e da codificação da profissão dos imigrantes

Codificação Profimg * País União Europeia Crosstabulation

			País União Europeia		Total
			sim	Não	
Codificação Profimg	Empregado	Count	53	59	112
		Expected Count	36,4	75,6	112,0
		Std. Residual	2,8	-1,9	
	Estudante	Count	21	39	60
		Expected Count	19,5	40,5	60,0
		Std. Residual	,3	-,2	
	Doméstico	Count	10	83	93
		Expected Count	30,2	62,8	93,0
		Std. Residual	-3,7	2,6	
	Outros	Count	19	33	52
		Expected Count	16,9	35,1	52,0
		Std. Residual	,5	-,4	
Total	Count	103	214	317	
	Expected Count	103,0	214,0	317,0	

Observa-se no Quadro 26 uma tabela contingência que mostra a relação entre os países da União Europeia e a profissão dos imigrantes, os imigrantes não Europeus têm numero superior na profissão doméstico (83) do que os Europeus (10). Os que têm origem fora da União Europeia vem mais com ocupação de estudantes relativamente com os que vem da União europeia. Os imigrantes que vêm da Europa vêm mais para outros motivos quando comparamos com aquilo que se esperava.

Quadro 27. Estatística teste do qui-quadrado

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	31,826 ^a	3	<0,001
Likelihood Ratio	35,359	3	,000
Linear-by-Linear Association	12,391	1	,000
N of Valid Cases	317		

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 16,90.

Pelo teste do qui-quadrado obtêm-se um valor $p = 0.000 < \alpha = 0.1$ pelo que existe evidência estatística para concluir que existe associação entre a profissão dos imigrantes e o seu país de origem, sendo que neste caso há muito mais imigrantes de fora da União Europeia de profissão doméstico. Existe evidência estatística de que a profissão dos imigrantes depende dos países de origem.

3.4. Teste T para amostras emparelhadas

O teste t permite inferir sobre a igualdade de médias de duas populações emparelhadas. Este teste é também apropriado para analisar dois grupos diferentes de casos relativamente a alguma característica que tenham em comum e pela qual possam ser comparados. Se não existir correlação entre os dois grupos ou se for pequena, significa que o emparelhamento não foi útil, devendo em consequência usar o teste t para amostras independentes, porque possui um maior número de graus de liberdade. Nota-se que deve haver sempre correlação entre dois grupos para se utilizar este teste. Como neste estudo só existe duas variáveis quantitativas então pode-se aplicar o teste t a comparar a variável idade do cônjuge (I_{conj}) e coma dos imigrantes (I_{imig}).

Portanto testar as hipóteses:

$$H_0 : \mu_{I_{conj}} \leq \mu_{I_{imig}}$$

vs

$$H_1 : \exists \mu_{I_{cjg}} > \mu_{I_{img}}$$

A estatística a utilizar é:

$$t = \frac{\bar{d} - \mu_d}{s'_d / \sqrt{n}} \sim t_{(n-1)};$$

com

$$\bar{d} = \frac{\sum d_i}{n}; s'_d = \frac{\sum (d_i - \bar{d})^2}{n-1}$$

Onde a média de d é \bar{d} e a variância de d é s'_d .

Em seguida podemos verificar o pressuposto básico a normalidade.

❖ Normalidade das distribuições

Para testar a normalidade podemos observar o quadro 28, como tamanho de $n = 354$ com $\alpha = 0,1$. Como a dimensão da amostra grande o teste a utilizar para testar a normalidade das distribuições é Kolmogorov-Smirnov. Para este teste, observa-se o valor p é inferior a $\alpha = 0,1$ pelo que rejeita-se a hipótese nula, segundo a qual não existe a normalidade da distribuição a idade dos imigrantes e dos seus cônjuges.

Quadro 28. Teste a normalidade

	Tests of Normality					
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Diferença Idade dos Cônjuges e dos imigrantes	,123	222	,000	,932	222	< 0,001

a. Lilliefors Significance Correction

Como a distribuição não é normal podemos então concluir estes dados não tem condições para aplicar o teste t e podemos passar o teste de não paramétrico é o teste de Wilcoxon. O teste de Wilcoxon aplica-se para analisar diferenças entre duas condições no mesmo grupo de sujeitos, em variáveis contínuas.

A estatística teste quando existem empates é dada por:

$$T = \frac{W^+ - W^-}{\sqrt{\sum_{i=1}^n R_i^2}}$$

Onde:

W^+ é a soma das ordens que correspondem a diferenças D_i positivas

W^- é a soma das ordens que correspondem a diferenças D_i negativas.

R_i é a ordem atribuída $|D_i|$.

$|D_i| = |Y_i - X_i|$ são dispostos por ordem crescente, atribuindo-se a ordem 1 à menor diferença, a ordem 2 à seguinte, e assim sucessivamente.

Em seguida vamos ver o resultado da análise;

Quadro 29. Ranks

		Ranks		
		N	Mean Rank	Sum of Ranks
Idade do conjugue - Idade dos imigrantes	Negative Ranks	73 ^a	80,90	5906,00
	Positive Ranks	130 ^b	113,85	14800,00
	Ties	18 ^c		
	Total	221		

a. Idade do conjugue < Idade dos imigrantes

b. Idade do conjugue > Idade dos imigrantes

c. Idade do conjugue = Idade dos imigrantes

No Quadro 29 apresentam-se os números de ordens negativas, positivas e empates (N). A média das ordens (Mean Rank) e a soma das ordens (Sum of Ranks) . existe a observação com diferença positiva a que corresponde a menor soma das ordens ($\min(S^+, S^- = 14800)$).

Quadro 30. Estatística de teste

Test Statistics ^b	
	Idade do conjugue - Idade dos imigrantes
Z	-5,311 ^a
Asymp. Sig. (2-tailed)	<0,001

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

O valor p exacto do teste unilateral pode obter-se a partir do teste bilateral dividindo-o por dois, isto é < 0.0001 , o que leva á rejeição de H_0 . Pode concluir-se que a idade dos cônjuges é maior do que idade dos imigrantes, ou seja maioria dos imigrantes procuram as pessoas mais velhas para garantir a sua estabilidade.

CAPÍTULO IV

Estudo Estatístico do Casamento dos Imigrantes

4.1. Modelação não Paramétrica

Analisar o tempo de vida é um assunto extremamente interessante para pesquisadores das áreas de biologia, medicina, engenharia, estatística entre outras. Neste caso estudámos o tempo de duração do casamento, o que é um exemplo na área de demografia.

A característica principal desse tipo de dados é a presença de censura ou eventos nas observações e isto conduz a algumas complicações de ordem técnica. Existem duas abordagens estatísticas para estimar a curva de sobrevivência, uma delas considera modelos não paramétricos e outras modelos paramétricos. A escolha de uma das abordagens não significa a exclusão da outra, por exemplo, o modelo não paramétrico pode indicar evidências de que um modelo paramétrico não está adequado. Por isso o uso das duas abordagens é essencial para garantir resultados verdadeiros.

A estimação da probabilidade de sobrevivência para diferentes horizontes de tempo permite obter o “perfil de sobrevivência” do casamento dos imigrantes na sua população. Além disso, esse modelo possui a vantagem de não estabelecer qualquer hipótese sobre a forma de distribuição dos dados.

Para a análise de sobrevivência, foram considerados como evento os divórcios que ocorreram nos casamentos dos/das imigrantes com portugueses/as. Foi definido como o tempo de início de casamento para cada indivíduo (T_0), a data em que foi celebrado o casamento em Portugal. O período do estudo teve início no primeiro dia de 2006 e fim a trinta e um de Dezembro de 2009.

Tabela 22. Sumários dos dados

Sexo		Categoria das idades dos imigrantes			
Feminino	Masculino	se < Q ₁	se Q ₁ -Q ₂	se Q ₂ -Q ₃	Se > Q ₃
142	23	1	46	40	78
Naturalidade dos Cónjuge		Nível Escolaridade			
Portugal	Outros	Básico	Secundário	Superior	
150	14	45	91	24	
União Europeia		Categorias Existência de filhos de outro casamento			
Sim	Não	Não Têm filhos		Têm filhos	
9	156	164		1	
Designação da morada		Categorias Existência de filhos de outro casamento			
Rural	Urbano	Não Têm filhos		Têm filhos	
79	86	164		1	
Celebração do casamento		Motivo de entrada			
Civil	União de Facto	Trabalho	Reagrupamento Familiar	Outro	
133	31	45	91	12	
Profissão dos Cónjuges		Nacionalidade dos imigrantes			
Empregado	Estudante	Brasileira	Africana	Europeus	Restantes
146	8	123	13	5	24

4.1.1. Curvas de Kaplan-Meier

Nesta secção vamos estimar as curvas de sobrevivência de Kaplan-Meier para as variáveis utilizadas neste estudo.

Com o propósito de fazer uma investigação preliminar dos dados na forma de uma análise univariada sobre o tempo até à ocorrência do evento, começamos por utilizar

métodos não paramétricos para estimar e comparar curvas de sobrevivência. Na estimação não paramétrica não aplicamos um modelo teórico para as falhas e/ou censuras observadas.

Os factores em estudo foram considerados como variáveis independentes. Numa avaliação exploratória inicial foram constituídas curvas de sobrevivência para cada variável independente. O método utilizado foi o estimador produto limite de Kaplan-Meier. Foram analisados como factores independentes as variáveis: sexo, idadecat, celebração do casamento, União Europeia, motivo de entrada, nível escolaridade, designação da morada, profissão dos imigrantes e dos cônjuges, nacionalidade dos imigrantes, números dos filhos do casamento, e números dos filhos do outro casamento. Considera-se que o tempo de sobrevivência, variável dependente ou de interesse, é definida como o tempo transcorrido entre o início da observação até que o evento ocorra.

❖ Comparação das curvas de sobrevivência

Tabela 23. Valores p dos testes de Logrank e Peto para a comparação das curvas de sobrevivência para as diferentes variáveis em estudo.

Covariável	Testes (Valor P)	
	Logrank	Peto
Sexo	0.6620	0.6530
Celebração do casamento	0.0576	0.0622
União Europeia	0.0016	0.0038
Motivo de entrada	0.0061	0.0065
Nível Escolaridade	0.0188	0.0158
Designação da Morada	0.7860	0.7250
Profissão dos imigrantes	0.7190	0.7220
Profissão do cônjuge	0.6470	0.7460
Nacionalidade dos imigrantes	0.0020	0.0052
Idade dos imigrantes codificados	0.5690	0.6180
Existência de filhos do casamento	0.0439	0.0390
Existência de filhos de outro casamento	0.5660	0.5680

Na tabela 23 são apresentados os valores p dos testes Log rank e Peto. Podemos verificar que as variáveis que foram significativas são celebração do casamento, União Europeia (País de origem), motivo de entrada, nível escolaridade, Nacionalidade dos imigrantes e existência de filhos do casamento para a alfa (α) 10%. Como na análise sobrevivência podemos apresentar o tempo médio, o tempo mediano estimado e os intervalos da confiança para as variáveis significativas.

❖ Sexo

Observa-se na figura 28 o gráfico de estimador Kaplan-Meier da função sobrevivência por sexo, tendo-se verificado um valor $p = 0.6530$ logo não se rejeita a hipótese de não haver diferença entre as curvas de sobrevivência dos grupos do sexo masculino e do sexo feminino, pelo que o sexo não é uma variável significativa.

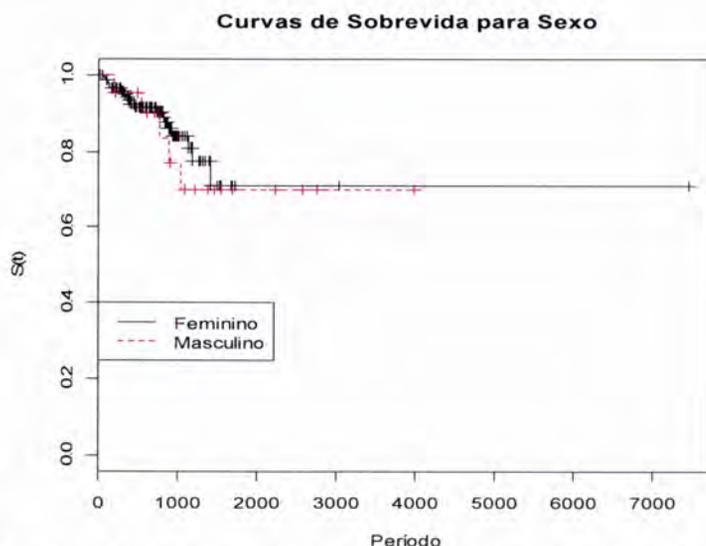


Figura 28. Curva do estimador Kaplan-Meier da função de sobrevivência para sexo

❖ Tipo de Celebração do casamento

Como podemos observar o valor p associado foi de $0.0576 < \alpha = 0.1$, este valor permiti-nos para rejeitar a hipótese nula, isto é, existe diferença estatisticamente significativa na sobrevivência entre os tipos de celebração do casamento. Isto, apesar de se observar que a curva de sobrevivência não se cruzam, no caso da união de facto está acima do que o tipo de celebração casamento de civil, o que no caso de ser significativo se poderia interpretar como os casamentos de união de facto durar mais que os civis (ver na figura 29).

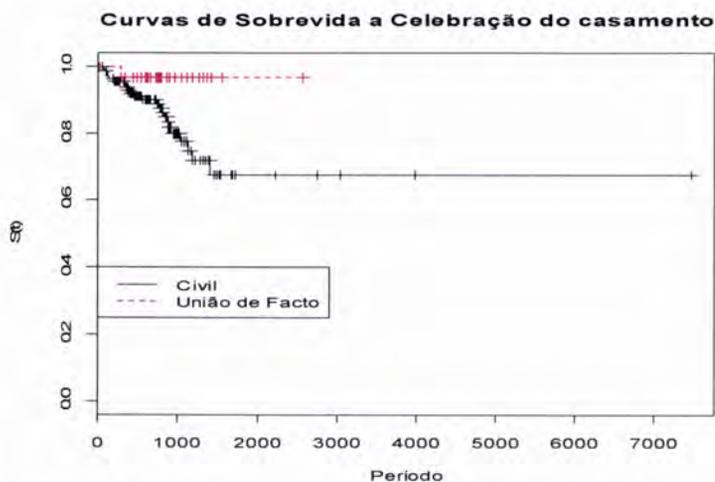


Figura 29. Curvado estimador Kaplan-Meier da função de sobrevivência para tipo de celebração do casamento

❖ União Europeia

Observa-se na Figura 30 que as curvas de sobrevivência para variável união europeia cruzam-se, por isso utiliza-se o peto. Como o valor de associado $P = 0.0038$ é menor do que $\alpha = 0,1$ rejeita-se a hipótese nula de igualdade entre a união europeia e não europeus quanto ao tempo de duração do casamento. Em particular, nota-se o tempo de duração do casamento dos imigrantes que fazem parte da União Europeia têm maior probabilidade do que os imigrantes que não fazem parte da União Europeia até 900 dias (onde as curvas se cruzam). Mas a depois dos 900 dias o tempo de duração de casamento para os imigrantes que não fazem parte da União Europeia tem maior probabilidade do que os imigrantes que fazem parte da União Europeia.

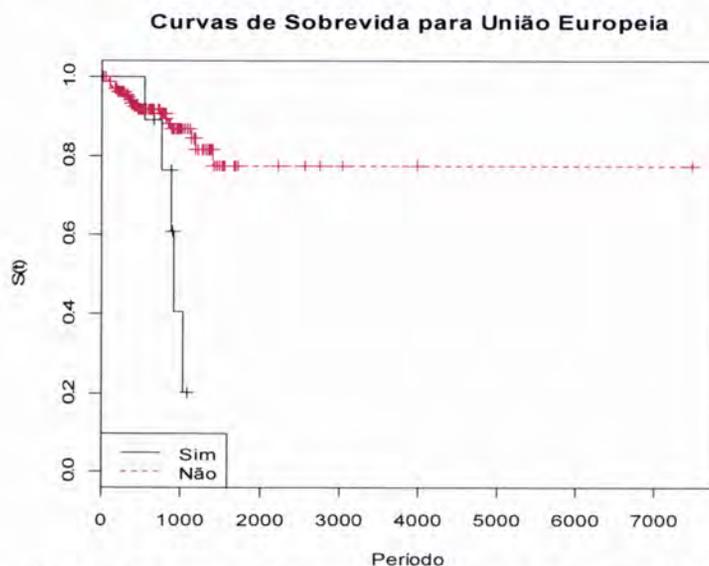


Figura 30. Curva do estimador Kaplan-Meier da função de sobrevivência para país de origem

No Quadro 31 podemos observar as estimativas que foram possíveis calcular para o tempo médio e tempo mediano de duração do casamento dos imigrantes nas duas categorias desta variável. Podemos observar que para os imigrantes de origem União Europeia a duração média do casamento é inferior a 3 anos (906 dias), sendo a mediana um valor muito próximo (917 dias), o que significa que metade dos casamentos envolvendo imigrantes da União Europeia não chega a durar 3 anos. Repare-se, através do Quadro 32, que $\frac{3}{4}$ dos casamentos têm uma duração que não excede 1030 dias (inferior a 3 anos, portanto). Já o casamento envolvendo imigrantes fora da União Europeia tem uma duração média superior a 16 anos o que é de realçar. Tal como em casos anteriores, só é possível estimar os quartis para os indivíduos que fazem parte da União Europeia, pois não há eventos depois do valor 75% da função de sobrevivência.

Quadro 31. Tempo médio e mediana da sobrevivência para a covariável país de origem (País de União Europeia)

Means and Medians for Survival Time

País União Europeia	Mean ^a				Median			
	Estimate	Std. Error	95% Confidence Interval		Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound			Lower Bound	Upper Bound
sim	906,359	58,801	791,108	1021,609	917,000	36,313	845,827	988,173
Não	5941,167	386,454	5183,718	6698,616
Overall	5610,303	394,533	4837,018	6383,588

a. Estimation is limited to the largest survival time if it is censored.

Quadro 32. Percentis para a covariável país de origem(País União Europeia)

Percentiles

País União Europeia	25,0%		50,0%		75,0%	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
sim	1030,000	.	917,000	36,313	881,000	219,681
Não
Overall	1406,000	.

❖ Motivo de Entrada

O comportamento das curvas mostra que o tempo de duração do casamento para os imigrantes que têm como motivo o reagrupamento familiar é maior do que os que têm motivo de trabalho e os outros.

Por outro lado, as curvas dos que têm motivos de trabalho e outros cruzam-se nos dias 950 e 1000, o que significa que nesse período o tempo de duração do casamento para os imigrantes que têm motivo de trabalho tem maior probabilidade do que os outros. O gráfico do outro motivo têm probabilidade de divorciar inferior do que os que tem motivo de trabalho e reagrupamento familiar. Observamos que as curvas cruzam-se, assim utiliza-se o teste peto, com o valor de p obtidas é $0.0065 < \alpha = 0.1$, rejeita-se a hipótese da igualdade dos grupos de motivo de entrada (ver na Tabela 23 e na Figura 31).

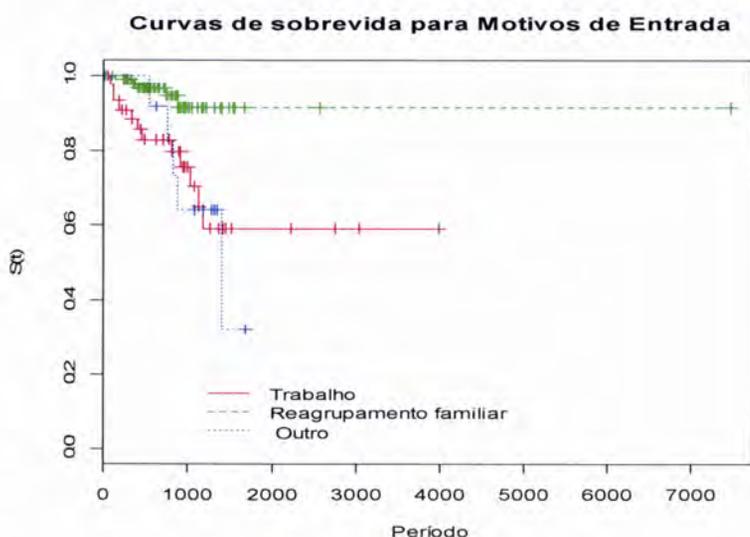


Figura 31. Curva do estimador Kaplan-Meier da função de sobrevivência para Motivo de entrada

O tempo médio de duração do casamento para os imigrantes que têm motivo de trabalho é de aproximadamente 2647 dias (pouco mais de 7 anos), sendo de 6854 dias (quase 19 anos) para os que entram por motivos de reagrupamento familiar e de 1264 dias (três anos e meio), para os que entram por outro motivo. No Quadro 33 podem ver-se as estimativas da média e da mediana e respectivos intervalos de confiança (obtidos sempre que possível) para as diferentes categorias desta variável, enquanto que no Quadro 34 se podem observar as estimativas dos quartis.

Quadro 33. Tempo médio e mediana da sobrevivência para a covariável motivo de entrada

Means and Medians for Survival Time

Codificação Moterna	Mean ^a				Median			
	Estimate	Std. Error	95% Confidence Interval		Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound			Lower Bound	Upper Bound
Trabalho	2647,412	320,849	2018,548	3276,276
Reagrupamento Familiar	6854,338	280,264	6305,021	7403,655
Outro	1263,654	132,145	1004,650	1522,658	1406,000	397,665	626,577	2185,423
Overall	5551,315	421,488	4725,198	6377,431

a. Estimation is limited to the largest survival time if it is censored.

Quadro 34. Percentis para a covariável motivo de entrada

Percentiles

Codificação Moterna	25,0%		50,0%		75,0%	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Trabalho	1030,000	318,737
Reagrupamento Familiar
Outro	.	.	1406,000	397,665	832,000	80,136
Overall	1406,000	.

❖ **Nível Escolaridade**

Observa-se na figura 32 que o tempo de duração do casamento ao nível de escolaridade secundário parece ser maior do que o básico e do superior. O tempo de duração do casamento do nível superior tem maior probabilidade do que o básico, ou seja os imigrantes que tem nível escolaridade básico tem probabilidade de sobreviver inferior aos outros. Como as curvas não se cruzam, utiliza-se o teste Logrank, verifica-se na Tabela 23 o valor $p = 0.0188 < \alpha = 0.1$, rejeita-se a hipótese da igualdade das curvas sobrevivência, ou

seja existe diferença significativa das curvas sobrevivência entre os grupos do nível de escolaridade básico, secundário e superior.

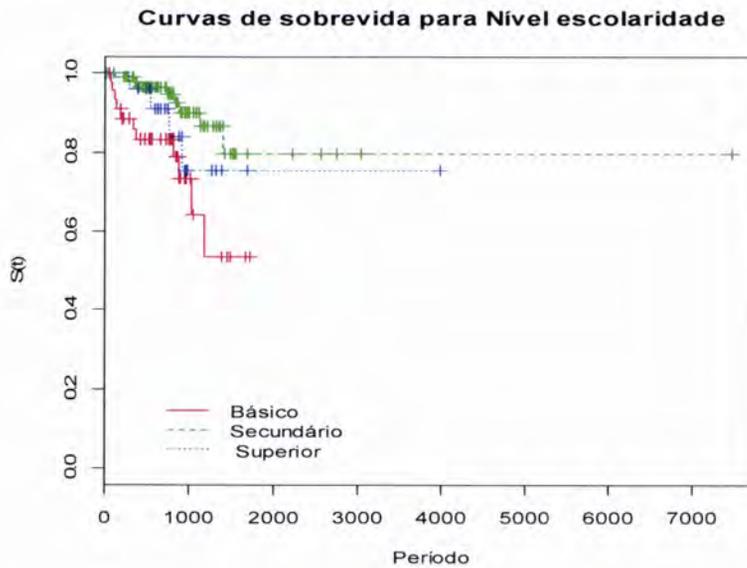


Figura 32. Curva do estimador Kaplan-Meier da função de sobrevivência para Nível Escolaridade

O tempo médio de duração do casamento para os imigrantes que têm nível escolaridade básico é de 1255 dias (três anos e meio) de 3175 dias (um pouco mais de 8 anos e meio) para os com nível de escolaridade superior e quase o dobro para os que têm nível de escolaridade secundário, 6137 dias (quase 17 anos). Não foi possível calcular estimativa para a mediana por categorias, bem como para vários dos quartis em virtude de termos muitas observações censuradas se o último evento permitir apenas a estimativa de um dos quartis e para uma categoria apenas. Deixamos os valores que conseguimos obter nos Quadros 35 e 36.

Quadro 35. Tempo médio e mediana da sobrevivência para a covariável nível escolaridade

Means and Medians for Survival Time

Codificação Nivescola	Mean ^a				Median			
	Estimate	Std. Error	95% Confidence Interval		Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound			Lower Bound	Upper Bound
Básico	1255,339	114,832	1030,269	1480,409
Secundário	6137,756	503,549	5150,800	7124,712
Superior	3175,130	365,514	2458,723	3891,538
Overall	5622,641	397,489	4843,562	6401,720

a. Estimation is limited to the largest survival time if it is censored.

Quadro 36. Os percentis para a covariável nível escolaridade

Percentiles

Codificação Nivescola	25,0%		50,0%		75,0%	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Básico	870,000	288,367
Secundário
Superior
Overall	1406,000	.

Em seguida apresenta-se as curvas de Kaplan-Meier e a sua interpretação das comparações das curvas de sobrevivência.

❖ Designação da Morada

Observa-se na Figura 33 o gráfico de estimador kaplan-Meier da função sobrevivência por designação da morada, do modo como as curvas se cruzam podemos dizer que as curvas são idênticas e acumulando. Com base nesta observação do gráfico permiti-nos a dizer o tempo de duração do casamento dos imigrantes que moram em áreas rurais quase igual

como os imigrantes que moram em meio urbano. Utilizando o teste peto para os grupos de designação da morada, verifica-se na Tabela 23 o valor de p-value, $0.7250 > \alpha = 0.1$, logo não se rejeita a hipótese o que significa não existe diferença significativa entre as curvas de sobrevivência dos grupos do rural e dos urbanos.

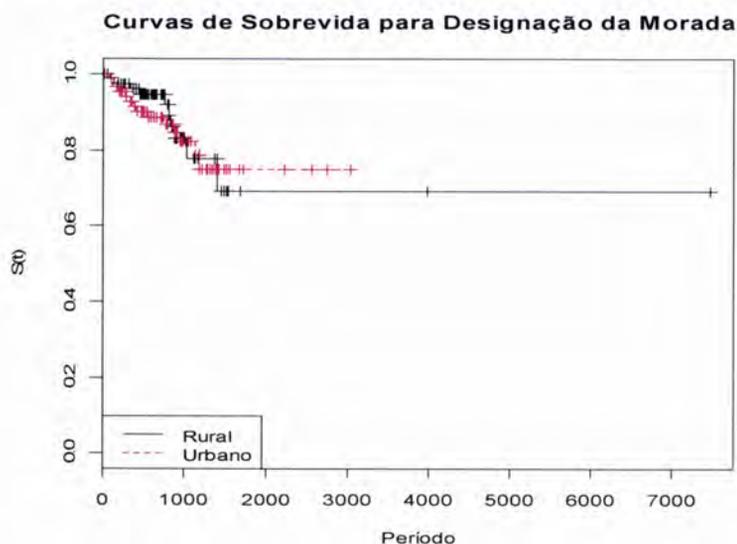


Figura 33. Curva do estimador Kaplan-Meier da função de sobrevivência para Nível Escolaridade

❖ Profissão dos imigrantes

Verifica-se na figura 34 que as curvas de sobrevivência para os grupos das profissões dos imigrantes cruzam-se nos empregados e dos outros, mas esses dois grupos não se cruzam com estudantes. Os que tem profissão de estudante, apesar de haver poucos indivíduos, mesmo assim ainda tem uma curva que fica acima das outras curvas. Por outro lado podemos ver na Tabela 23 que o valor de p-value = $0.7220 > \alpha = 0.1$, logo não se rejeita a hipótese da igualdade das curvas de sobrevivência entre os grupos de empregado, estudante e outro. Conclui-se que não existe diferença entre as curvas de sobrevivência dos grupos de profissão dos imigrantes, dos empregados, estudante e outro.

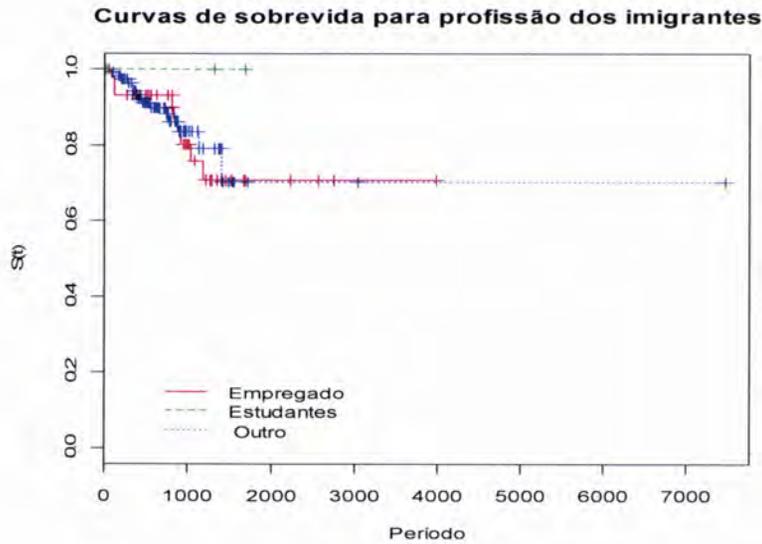


Figura 34. Curva do estimador Kaplan-Meier da função de sobrevivência para profissão dos imigrantes

❖ Profissão dos cônjuges

As curvas de sobrevivência para variável profissão dos cônjuges mostra que as outras profissões dos cônjuges têm maior probabilidade no período até próximo dos 1000 dias, e depois no início de 1000 dias verificou-se que o tempo de sobreviver para os empregados já maior do que para os outros, ou seja no início de período 1000 dias a probabilidade de sobreviver para as outras profissões dos cônjuges é inferior do que os empregados.

O valor de $p = 0.7460$ é maior do que $\alpha = 0.1$, logo não se rejeita a hipótese nula de igualdade entre o grupo de empregado e o grupo de outro (ver na Tabela 23 e figura 35). Concluimos então que as curvas de sobrevivência não diferem entre as profissões dos cônjuges para o empregado e da outra profissão.



Figura 35. Curva do estimador Kaplan-Meier da função de sobrevivência para profissão dos cônjuges

❖ Nacionalidade dos Imigrantes

Pela Figura 36, as curvas de sobrevivência estimadas para os grupos de codificação nacionalidades por classes cruzam-se. Outra informação é que os imigrantes com nacionalidade brasileiros, africanas, União Europeus, e os restantes cruzam ao mesmo tempo entre 900 e 1000 dias, e que acontece mais eventos para o grupo com nacionalidade brasileira. Parece que o tempo de duração de casamento para indivíduos da União Europeus é menor do que as outras nacionalidades, passando-se o inverso com os africanos.

Observa-se na tabela 23 o teste Peto, com o valor de $P\text{-value} = 0.0052 < \alpha = 0.05$, logo rejeita-se a hipótese nula de igualdade das curvas sobrevivência entre os diferentes grupo de nacionalidade codificados, o que significa que existe diferença significativa nas curvas sobrevivência entre os grupos dos brasileiros, africanos, união europeus e restantes nacionalidade codificados.

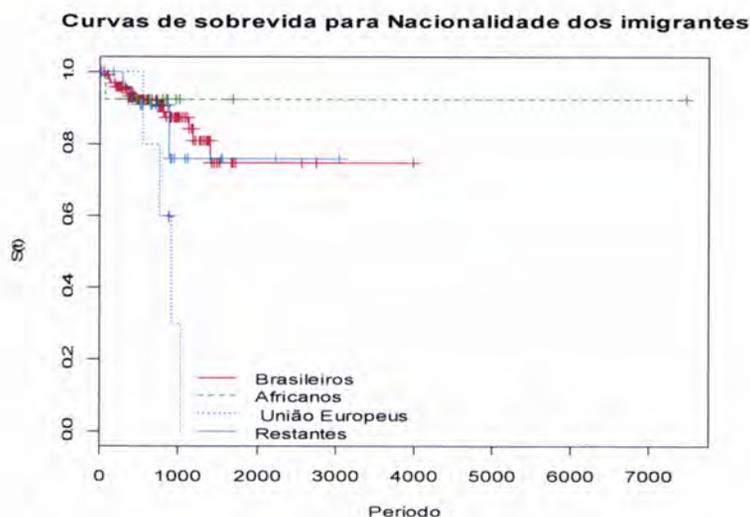


Figura 36. Curva do estimador Kaplan-Meier da função de sobrevivência para Nacionalidade dos imigrantes

❖ Idade dos imigrantes

Observando as curvas da idade dos imigrantes codificados em duas classes verifica-se que as curvas são praticamente coincidentes até aos 1000 dias, sendo que a partir dessa altura os casamentos de imigrantes com idade superior tem maior probabilidade de ser mais duradouro que o casamento de imigrantes mais novos. A coincidência das curvas de sobrevivência também se verifica até aos 1000 dias no caso em que tomamos 4 classes, sendo que a partir daí também se visualiza algo idêntico ao que se passa com duas classes. Contudo, quer num caso quer noutra a amostra não revelou evidência significativa para

concluirmos que existe diferença das curvas de sobrevivência e, conseqüentemente, a variável idade não explica por si o tempo de duração do casamento ver na Figura 37 e 38.

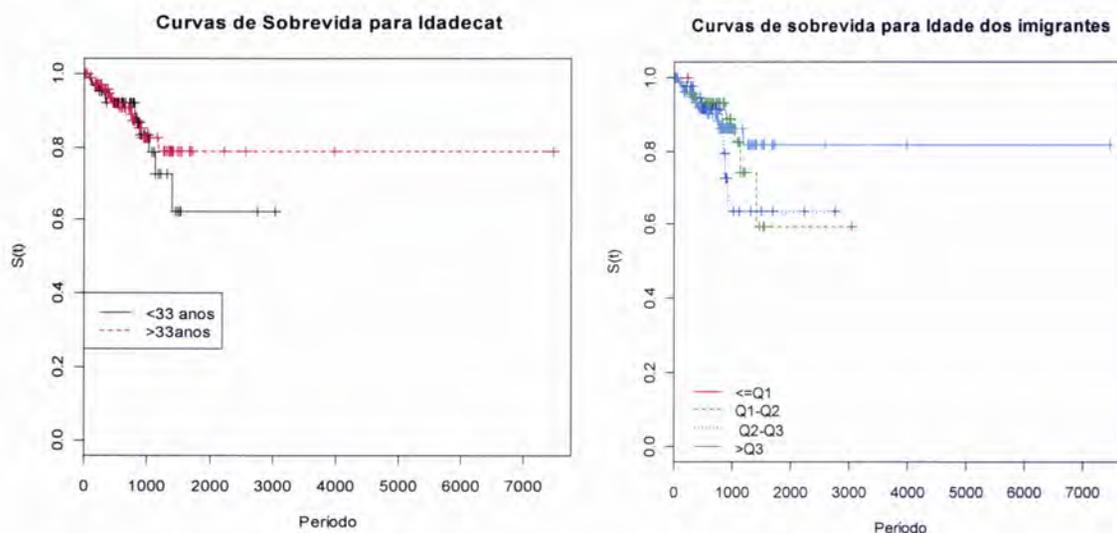


Figura 37. Curvas do estimador Kaplan-Meier da função de sobrevivência para idade dos imigrantes codificados por duas e por quatro classes.

❖ Existência de filhos do casamento e de outro casamento

Observa-se pelo gráfico na Figura 38, as curvas de sobrevivência estimadas não se cruzam para as duas variáveis, existência de filhos do casamento e existência de filhos de outro casamento, o que significa que o casamento dos que não têm filhos têm maior probabilidade de durar mais do que o casamento dos que têm filhos. Relativamente à variável existência de filhos de outro casamento, como os eventos foram muito poucos é natural que não se revelem diferenças significativas entre as duas categorias. Podemos, no entanto, observar que o período de maior risco de término do casamento nos que não têm filhos de outro casamento vai até aos 4 anos

Como as curvas não se cruzam utiliza-se o teste Logrank para a variável de existência de filhos do casamento o valor $p = 0.0439 < \alpha = 0.1$, logo rejeita-se a hipótese nula de igualdade das curvas sobrevivência entre os diferentes grupos de existência de filhos do casamento. Por outro lado, para a variável existência de filhos de outro casamento continua-se utilizar o teste Logrank que mostra o valor $p = 0.5660 > \alpha = 0.1$, portanto não se rejeita a hipótese nula, o que significa que não existe diferença significativa nas curvas sobrevivência entre os grupos dos que tem filhos e dos que não tem filhos.

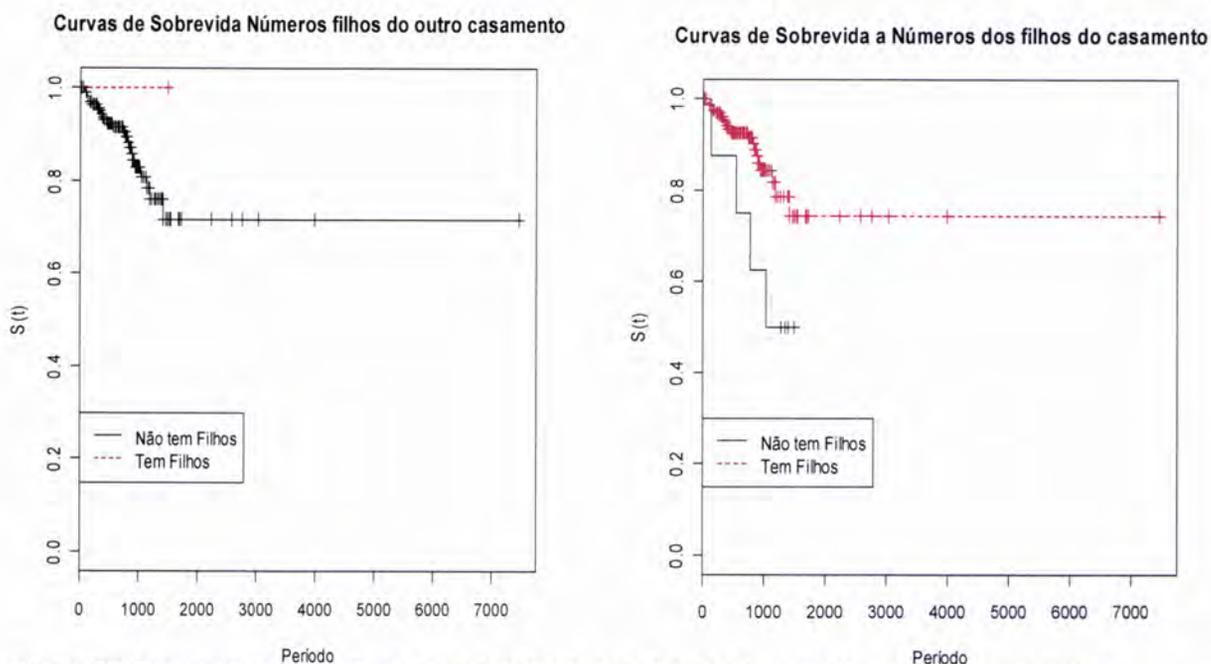


Figura 38. Curvas de sobrevivência para existência dos filhos do casamento e do outro casamento

4.1.2. Modelo de Regressão de Cox

Após a análise inicial, as variáveis que apresentaram significância estatística ou foram consideradas de importância, devem integrar o Modelo de Cox. Recorde-se que nesta análise estão incluídos 165 imigrantes que casaram com portugueses/as.

Aplicando o modelo de Cox univariadamente obtivemos os resultados que se encontram na Tabela 24.

Uma análise preliminar indica que, possivelmente as variáveis Sexo, Profissão dos imigrantes, profissão dos cônjuges, designação da morada, Naturalidade dos imigrantes, Nacionalidade dos imigrantes, idade dos imigrantes e dos cônjuges, não são importantes para explicar o tempo duração de casamento até acontecer o divórcio entre um/a imigrante que casou com um/a português/a. Porquê, essas variáveis não são significativas, nota-se nos valores do P-value são superiores do que $\alpha = 0,1$.

As covariáveis que foram significativas; celebração do casamento, país de origem (União Europeia), motivo de entrada, nível escolaridade, nacionalidade dos imigrantes, existência filhos do casamento. Nota-se essas covariáveis foram significativas em Kaplan-Meier.

Tabela 24. Regressão de Cox para todas as variáveis dos dados dos imigrantes

Variável	Beta (β)	HR= exp (β)	Valor P	IC _{95%} Exp β
SexoM	0.0222	1.2486	0.6630	[0.4602; 3.3880[
Nível Escolaridade 1	-1.2439	0.2882	0.0077	[0.1155; 0.7191[
Nível escolaridade 2	0,06008	0.5484	0.3000	[0.1744; 1.7246[
Profissão dos imigrantes 1	-1.710e+01	3.760e-08	0.9980	[0.0000; Inf[
Profissão dos imigrantes 2	-6.757e-02	9.347e-01	0.8740	[0.4044; 2.1600[
Profissão dos cônjuges	0.1692	1.1843	0.6490	[0.5722; 2.4510[
União Europeia	-1.4849	0.2265	0.0039	[0.0828; 0.6202[
Designação da Morada	0.1128	1.1194	0.7860	[0.4953; 2.5300[
Tipo de celebração do casamento	-1.7197	0.1791	0.0923	[0.0242; 1.3270[
Nacionalidade dos imigrantes1	-0.4269	0.6525	0.6796	[0.0861; 4.9470 [
Nacionalidade dos imigrantes2	1.8598	6.4228	0.0011	[2.0948; 19.6930[
Nacionalidade dos imigrantes3	0.2333	1.2628	0.6788	[0.4186; 3.810 [
Naturalidade dos cônjuges	-0.0827	0.9206	0.9110	[0.2150; 3.9430[
Nacionalidade dos cônjuges	0	1	NA	[1; 1[
Idade dos imigrantes	0.0231	1.0234	0.2700	[0.9822; 1.0660 [
Idade dos cônjuges	-0.0065	0.9936	0.7220	[0.9589; 1.0300 [
Idade dos imigrantes codificados 1	1.401e+01	1.217e+06	0.998	[0; Inf [
Idade dos imigrantes codificados 2	1.437e+01	1.735e+06	0.998	[0; Inf [
Idade dos imigrantes codificados 3	1.384e+01	1.020e+06	0.998	[0; Inf [
Motivo de entrada1	-1.4484	0.2349	0.0069	[0.08222;0.6713[
Motivo de entrada2	0.1583	1.1715	0.7673	[0.4115; 3.3356[
Existência de filhos do casamento	-1.0636	0.3452	0.0542	[0.1169; 1.0190 [
Existência de filhos do outro casamento	-16.0000	1.09e-07	1	[0; Inf [

A interpretação das covariáveis que significativas vão apresenta-se em seguintes:

❖ Nível escolaridade:

Os imigrantes que têm nível escolaridade secundário têm razão de risco 0.2882, significa que para os imigrantes que têm nível escolaridade secundário

o risco de divórcio diminui 71,18% quando comparados com os que têm nível escolaridade básico.

❖ Origem

Os imigrantes de origem dentro da União Europeia têm um risco 4 vezes superior aos imigrantes de origem fora da União Europeia.

❖ Tipo de celebração do casamento

Os imigrantes que foram celebrar o casamento por civil têm a razão de risco 0.1791, significa os imigrantes que viveram juntos ou União de facto tem o risco aumentar 6 vezes relativamente os que casaram por civil.

❖ Nacionalidade dos imigrantes

Os imigrantes Europeus têm a razão de risco 6.4228, significa os imigrantes de nacionalidade europeia tem o risco 6 vezes superior relativamente os brasileiros.

❖ Motivo de entrada

Os imigrantes que têm motivo de trabalho têm um risco 4 vezes superior relativamente aos que entram por motivo de reagrupamento familiar.

❖ Existência de filhos do casamento

Os imigrantes que têm filhos têm um risco 66% menor dos que não têm filhos, ou seja nos imigrantes que não têm filhos o risco aumenta 3 vezes.

A análise preliminar foi aplicar o modelo de Cox para todas as covariáveis individualmente, em seguida junta-se as covariáveis que foram significativas. Depois de entrar as covariáveis pode-se excluir uma a uma que não foi significativa. As covariáveis que foram significativas podemos ver a interacção, mesmo assim resultado da interacção não foram significativas. Por isso as covariáveis seleccionadas foram motivo de entrada, nível escolaridade e país de origem (União Europeia).

Tabela 25. Modelo final do modelo Cox

Modelo	Covariáveis	Estimativas Beta (β)	HR = exp (β)	Valor P	IC _{95%} Exp β	Log Verosimilhança
	Motivo reagrupamento familiar	-1,1627	0,3126	0,0402	[0.1030; 0.9491 [-108,3175
	Outro motivo	0,7623	2,1431	0,2010	[0.6663; 6.8930 [
	Nível secundário	-1,4846	0,2266	0,0025	[0.0866; 0.5926 [
	Nível superior	-1,6994	0,1828	0,0201	[0.0436; 0.7660 [
	Não União Europeia	-1,3497	0,2593	0,0182	[0.0846; 0.7948 [

O modelo de cox dada por:

$$h(t) = h_0(t) \exp\{-1,16Mot.RA + 0,76Mot.Out - 1,49Niv.Sec - 1,70Niv.Sup - 1,35N\grave{a}o.UE\}$$

Para perfis fixos (relativamente às covariáveis que estão no modelo) podemos tecer as seguintes considerações:

- ❖ Os imigrantes que têm motivo de reagrupamento familiar tem a razão de risco igual 0.3126, significa os imigrantes que têm motivo de reagrupamento familiar têm o risco diminuir 69% relativamente os que têm motivo de trabalho.
- ❖ Os imigrantes que têm outro motivo tem a razão de risco igual 2.1431, significa os imigrantes que têm outro motivo têm o risco aumentar 2 vezes superior relativamente os que têm motivo de trabalho.
- ❖ Os imigrantes que têm nível escolaridade secundário têm o risco diminuir 77% dos imigrantes que têm nível escolaridade básico, ou seja os imigrantes que têm nível escolaridade básico tem o risco aumentar 4 vezes relativamente os imigrantes que têm nível escolaridade secundário.
- ❖ Os imigrantes que têm nível escolaridade superior têm o razão de risco 0.1828, seja os imigrantes que têm nível escolaridade superior tem o risco diminuir 82% dos imigrantes que têm nível escolaridade básico. Podemos dizer que os imigrantes que têm nível escolaridade básico têm o risco

umentar aproximadamente 6 vezes relativamente os imigrantes que têm nível escolaridade superior.

- ❖ Os imigrantes de origem da União Europeia têm um risco aumenta 4 vezes superior aos imigrantes que de fora da União Europeia.

Traduzindo o modelo de Cox por interpretação, este solicita que a taxa de risco num determinado instante no tempo pode ser desdobrada numa taxa de risco basal $h_0(t)$ e uma taxa de risco que é determinada pelo valor das variáveis independentes. O aumento ou diminuição da taxa de risco é uma razão constante através dos valores de cada variável independente, e por esse motivo o efeito destas variáveis sobre a taxa de risco é exponencial. Também por este motivo o modelo de Cox é um modelo risco proporcionais.

Para avaliar a adequacidade geral do modelo de Cox ajustado, tem de verificar os pressupostos do risco proporcional. Os pressupostos do risco proporcional são seguintes: os gráficos do log-log, Resíduos schoenfeld vs tempo, resíduos martingal, observações influentes.

❖ Avaliação da Proporcionalidades dos Riscos

Para verificação da suposição de riscos proporcionais no modelo de Cox, são apresentados os gráficos são seguintes:

a. Gráfico Log-log para as covariáveis

Conforme observado na Figura 39, as curvas não indicam violação da suposição de riscos proporcionais para as covariáveis estado civil, nível escolaridade, motivo de entrada, país da união europeia, por apresentarem diferenças aproximadamente constantes ao longo do tempo. Observa-se que as curvas dos gráficos de log-log são aproximadamente paralelas o que significa que podemos admitir o pressuposto de riscos proporcionais

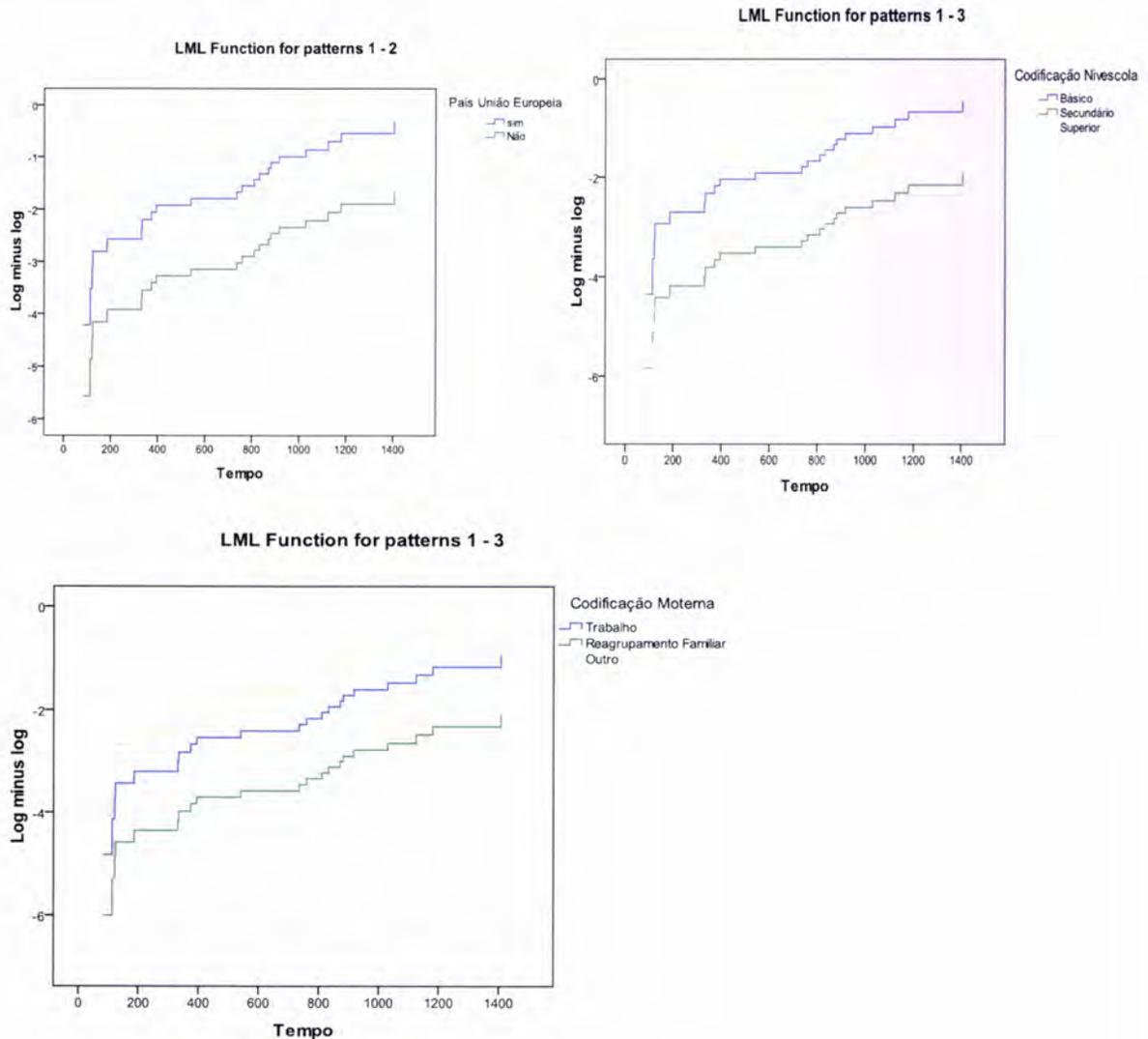


Figura 39: $\text{Log}(\hat{\Lambda}_{0j}(t))$ versus tempo para as covariáveis

b. Os resíduos de Schoenfeld versus o tempo de cada covariável

Para verificar o pressuposto de riscos proporcionais de resíduos de Schoenfeld versus o tempo podemos utilizar o gráfico e o teste Honell's.

Na figura 40, observa-se que para os resíduos do Schoenfeld as curvas são suaves e não é possível identificar tendências nas curvas, o que significa não haver evidência de violação da suposição de riscos proporcionais, ou seja com base nos resíduos do Schoenfeld os riscos são proporcionais. Por outro lado, realizando o teste de Honell's verifica-se que os valores p para as covariáveis motivo de entrada 1, motivo de entrada 2, nível escolaridade 1, nível escolaridade 2 são superiores do que alfa 5%, menos a covariável União Europeia, concluímos a maioria das covariáveis ajuste o modelo de risco proporcionais (ver na Tabela 26).

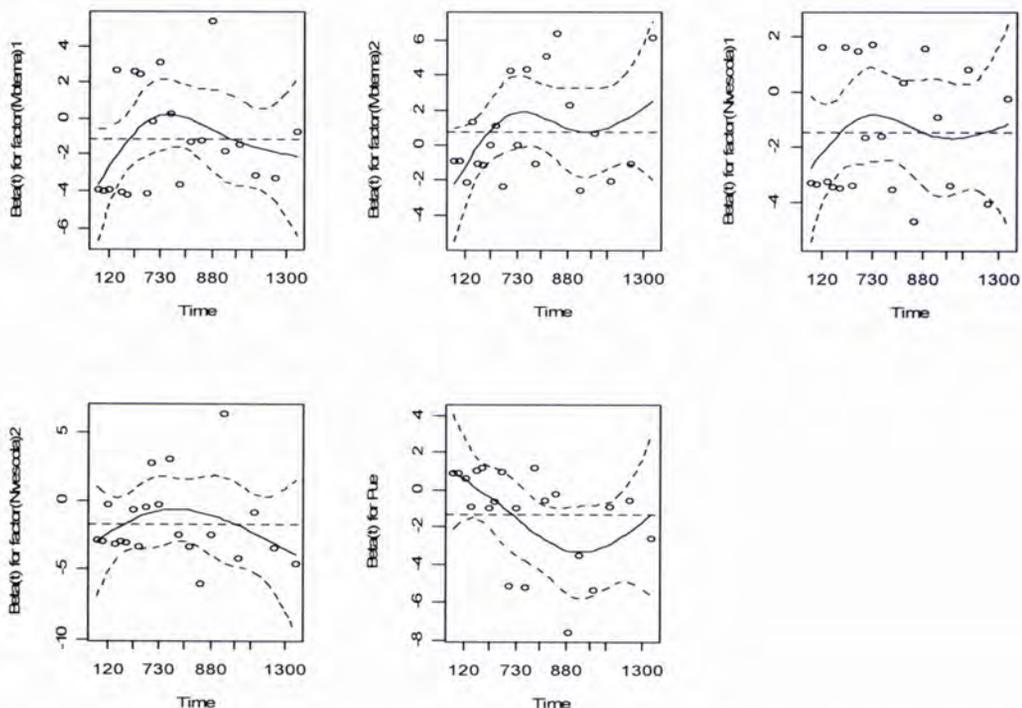


Figura 40. Resíduos de padronizados de Schoenfeld versus ao tempos para as covariáveis seleccionados

Tabela 26. Valores p do Teste de Honell's

Covariável	Rho (p)	X ²	Valor P
Motivo de entrada1	0.0701	0.1240	0.7248
Motivo de entrada2	0.2840	1.7740	0.1829
Nível escolaridade1	0.0844	0.1490	0.6997
Nível escolaridade2	-0,0711	0.0750	0.7842
União Europeia	-0.4424	3.8000	0.0513
Global	NA	5.3460	0.3751

c. Os gráficos dos resíduos martingal e Deviance do modelo ajustado

Observa-se na figura 41 os resíduos martingal e Deviance, podemos dizer que os resíduos de martingal surgem a existência de pontos que podem considerar como atípicos (Outliers). No resíduo Deviance o comportamento em torno de zero, assim indica-se esse resíduo é favorável à adequação do modelo ajustado aos dados deste estudo.

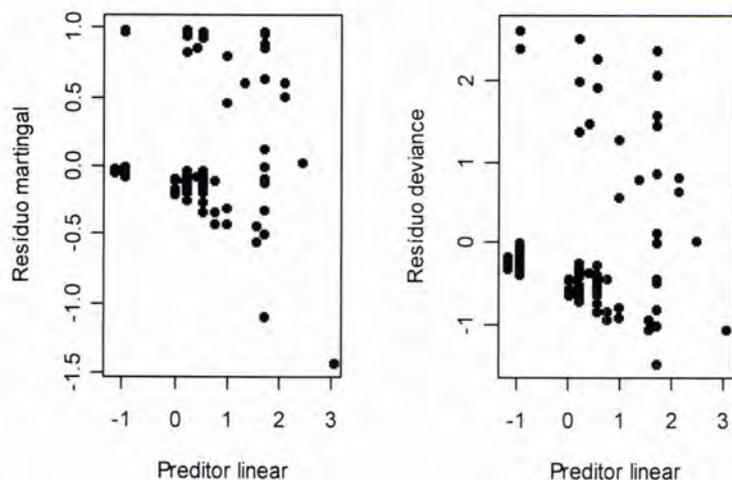


Figura 41. Resíduos martingal e Deviance versus predictor linear do modelo de Cox final ajustados para os dados de casamento

d. Resíduos Martingal

Através dos resíduos martingal podemos tentar identificar as observações influentes, ver a forma funcional se está correcto no modelo. A figura 42 mostra outra vez os resíduos martingal para todos os indivíduos, podendo identificar os pontos que se afastam dos outros, esse ponto é indivíduo 7 e 81. Mesmo assim com base nos gráficos podem dizer que os resíduos martingal verificam que os riscos são proporcionais porque não há uma tendência clara para os considerar como não aleatórios.

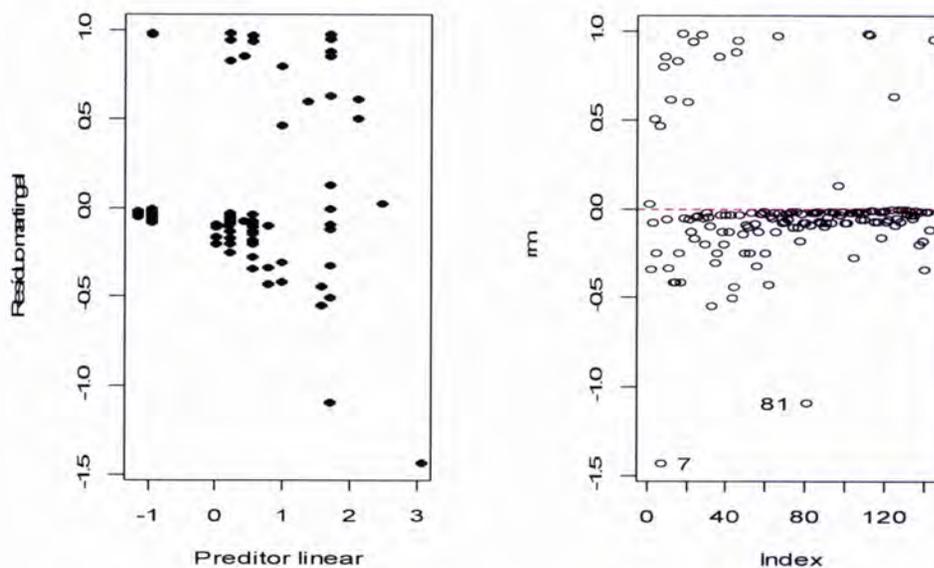


Figura 42. Resíduos martingal.

e. Pontos influentes

Como nos resíduos martingal foram identificar dois pontos foram indivíduo 7 e 81 mas verificou nos pontos influentes através deste resíduo só existe um ponto que fica muito afastado dos outros, este indivíduo é indivíduo 7 ver na Figura 43. Então, precisamos de ver se ele é influente ou não. Para a avaliação da influência de cada observação no ajustamento de um modelo utilizam-se os resíduos auxiliares ou resíduos dfbetas.

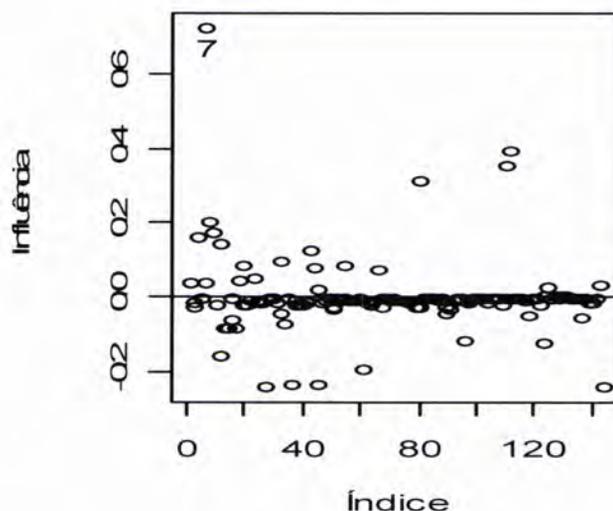


Figura 43. Resíduos dos pontos influentes

Os resíduos martingal mostra-se um ponto que fica afastar dos outros pontos, podemos considerar que este ponto é indivíduo 7 que considerou como ponto influente. A seguir pode-se tentar fazer análise com indivíduo 7 e sem indivíduo 7 na tabela 27 mostra-se existe grande diferença a 50 % por isso pode concluir-se que o ponto 7 é influente.

Tabela 27. Comparações dos valores dos coeficientes do modelo com e sem a observação 7

Covariáveis	Estimativas Beta (β) com indivíduo 7	Estimativas Beta (β) sem indivíduo 7	$\Delta\beta = \frac{\beta_{com7} - \beta_{sem7}}{\beta_{com7}} * 100$
Motivo de entrada1	-1.1627	-1,3040	-12,1351
Motivo de entrada2	0.7623	0,7420	2,6045
Nível escolaridade1	-1.4846	-1,8030	-21,4306
Nível escolaridade 2	-1.6994	-2,2830	-34,3301
União Europeia	-1.3497	-2,0310	-50,5104

Em seguida vamos comparar o risco para dos perfis onde as categorias dos perfis 1 e um imigrante que entre por motivo de trabalho tem o nível de escolaridade secundário com país de origem fora da União Europeia comparamos com o perfil 2. Um individuo com perfil 2 composto por seguinte características têm o nível escolaridade superior, veio de União Europeia e entre pelo mesmo motivo. Assim podemos categorizar os perfis e calcular o risco de perfil 1 comparando com perfil 2.

Tabela 28. Codificação dos perfis

Perfil 1	codificação	Perfil 2	codificação
Secundário	1	Superior	1
União Europeia	0	União Europeia	1
Trabalho	0	Trabalho	0

Para calcular o risco de $h(t)$ tomamos:

$$h(t) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)$$

$$HR = \frac{h_1(t)}{h_2(t)}$$

$$h_1(t) = h_0(t) \exp(-1,16 * 0 + 0,76 * 0 - 1,49 * 1 - 1,70 * 0 - 1,35 * 0)$$

$$h_1(t) = h_0(t) \exp(-1,49)$$

e

$$h_2(t) = h_0(t) \exp(-1,16 * 0 + 0,76 * 0 - 1,49 * 0 - 1,70 * 1 - 1,35 * 1)$$

$$h_2(t) = h_0(t) \exp(-1,15 - 1,91)$$

$$h_2(t) = h_0(t) \exp(-3,05)$$

Portanto para calcular hazard ratio ou função de risco é

$$HR = \frac{\exp(-1,49)}{\exp(-3,05)}$$

$$HR = \frac{0,225373}{0,047359}$$

$$HR = 4,76$$

Observa-se o valor de razão de risco para um indivíduo de perfil 1 é 4.76, significa um imigrante que entre por motivo de trabalho e seja da União Europeia com ensino secundário tem o risco aproximadamente 5 vezes superior de ter separado ou divorciado do que um imigrante que entra pelo mesmo motivo, vindo de um país fora da União Europeia com ensino superior.

4.1.3. Modelos Paramétricos

Apesar de existirem técnicas não paramétricas que tratam de dados sobre análise de sobrevivência, também há a necessidade de ajustar os dados ao modelo paramétrico mais apropriado. Escolha o modelo paramétrico para analisar estes dados porque este modelo também é importante a analisar tempo de vida. Por outro lado como o modelo não paramétrico tem poucas covariáveis significativas então podemos aplicar a análise paramétrica para ver se alguns modelos encaixados melhor dar algumas importâncias. Embora exista uma série de modelos probabilísticos em análise de sobrevivência, alguns destes ocupam posição de destaque por sua comprovada adequação a várias situações práticas, sendo usados com bastante frequência. Estes são o modelo exponencial, Weibull, Log-normal, e Log-logístico.

❖ Estimação dos parâmetros do modelo

a. Linearização para Exponencial, Weibull e Log-normal

Observa-se na Figura 44 os gráficos dos três modelos mostram-se um afastamento e os pontos não se aproximam bem de uma recta. Por isso a linearização não é adequada para modelos paramétricos.

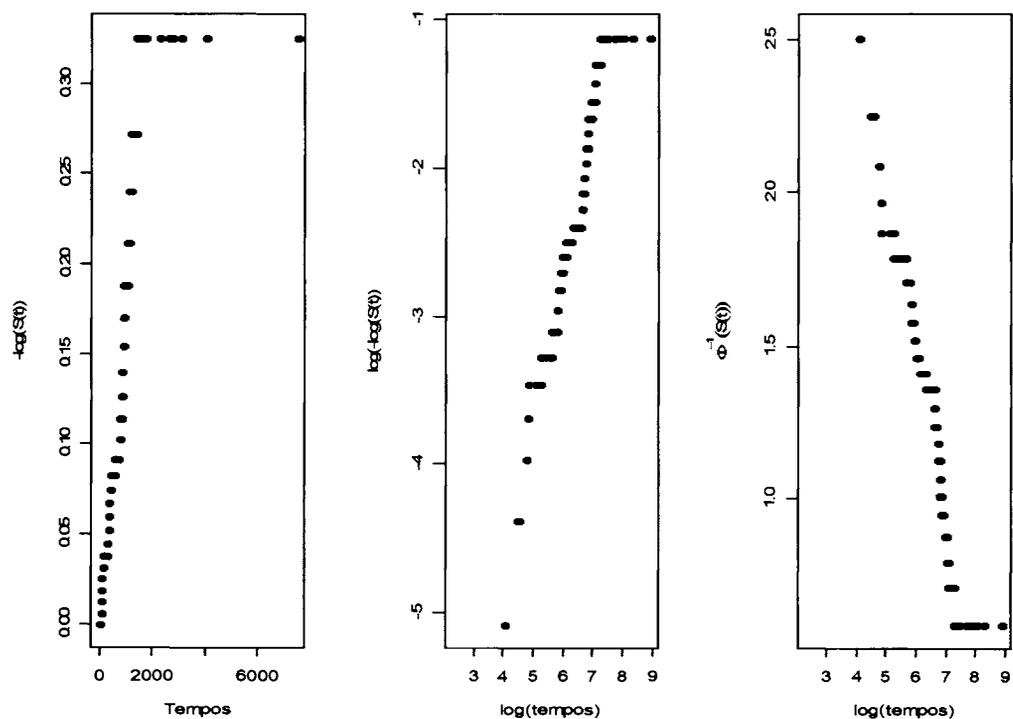


Figura 44. Gráficos de t versus $-\log(\hat{S}(t))$, $\log(t)$ versus $\log(\hat{S}(t))$ e $\log(t)$ versus $\Phi^{-1}(\hat{S}(t))$

b. Estimativas de Kaplan-Meier vs Estimativas dos modelos paramétrico

Ao compararmos as estimativas das funções de sobrevivência de cada um dos modelos em estudo com as estimativas de Kaplan-Meier obtemos os gráficos da Figura 60.

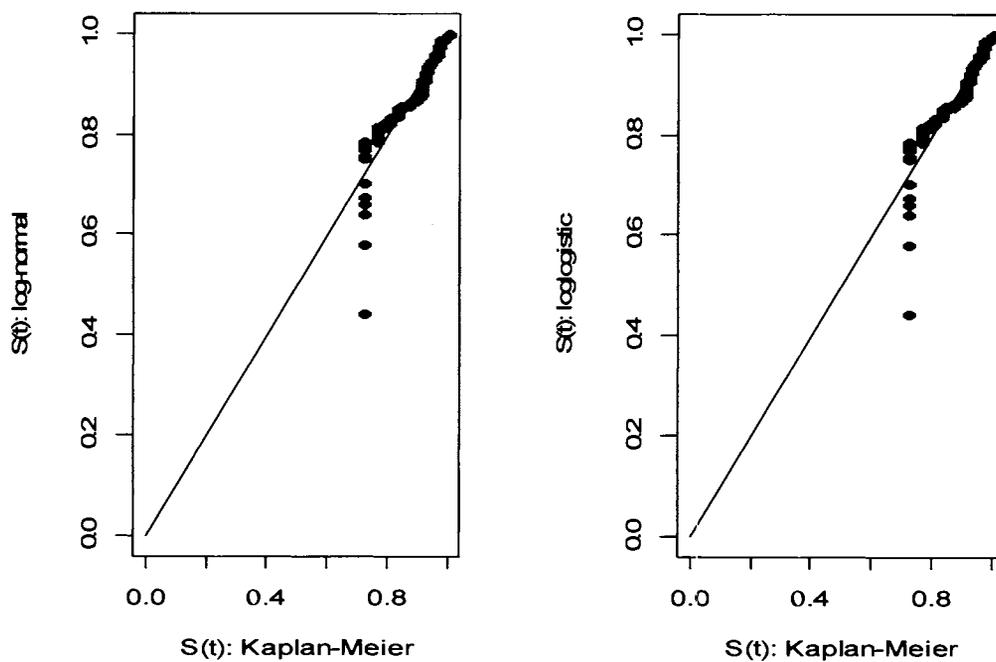


Figura 45. Gráficos das sobrevivências estimadas, por Kaplan-Meier versus as sobrevivências estimadas pelos modelos Log-normal e Log-logístico

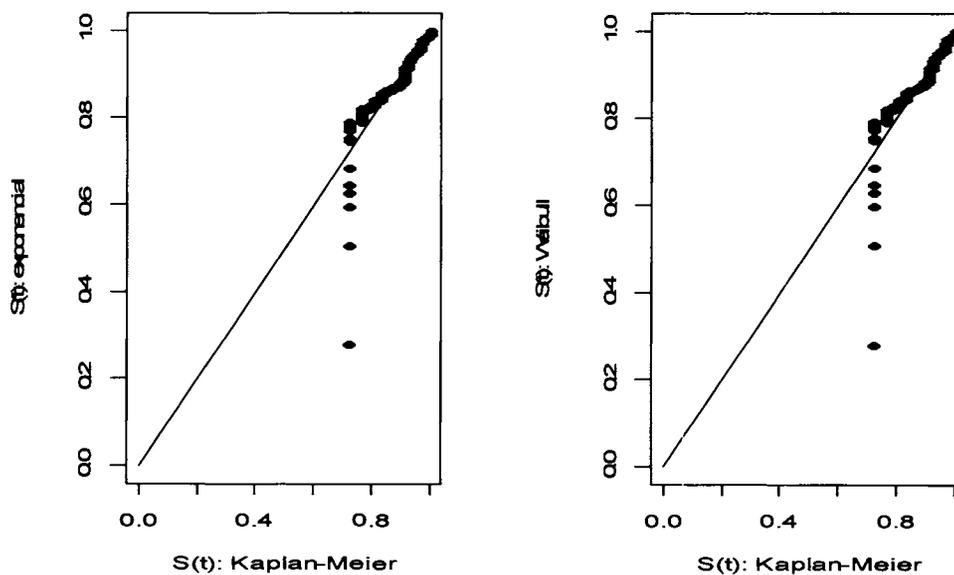


Figura 46. Gráficos das sobrevivências estimadas por Kaplan-Meier versus as sobrevivências estimadas pelos modelos Exponencial e Weibull

Podemos verificar que vários pontos se afastam bastante da recta pelo que nenhum destes modelos paramétricos parece adequado. O mesmo se pode comprovar nos gráficos das Figuras 45 e 46.

c. Análise Deviance e AIC

A partir desses resultados é possível concluir que os modelos não são adequados, pois o valor p é superior a $\alpha = 0.1$ isto significa que estes quatro modelos não são adequados para análise paramétrica. Por outro lado observa-se o valor de AIC são muitos grandes como os pressupostos do modelo não foi verificado ou seja não adequados por isso o modelo paramétrica não são possíveis para aplicar este dados

Tabela 29. Logaritmo da função $L(\theta)$ e AIC

Modelo	Log(L(θ))	Valor P	AIC
Exponencial	-232,6674	0,798	467.3348
Weibull	-232,6658	0,794	469.3317
Log-normal	-231,6677	0,153	467.3541
Log-Logística	-232,11	0,643	468.2234

Tabela 30. Teste Deviance

Modelo	Term Resid	Df	-2*LL Test	Df	Deviance	P(> Chi)
1	1	164	465,3348	NA	NA	NA
2	1	163	465,3317	1	0,0031	0,9555
3	1	163	463,3541	0	1,9775	NA
4	1	163	464,22	0	-0,8659	NA

d. Comparação das curvas de sobrevivência dos modelos com as curvas de sobrevivência de KM

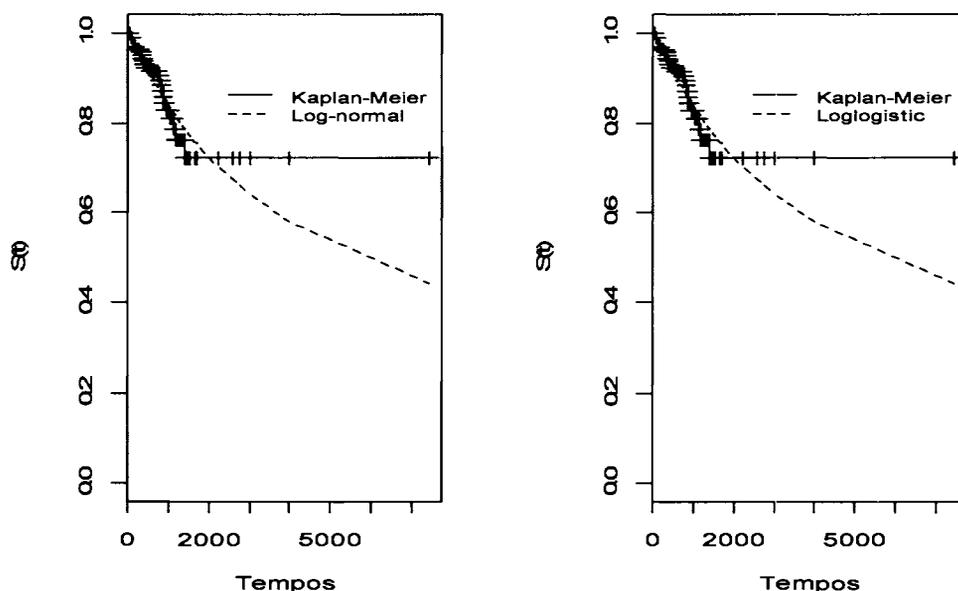


Figura 47. Curvas de sobrevivências estimadas pelos modelos de Log-normal e Log-logístico

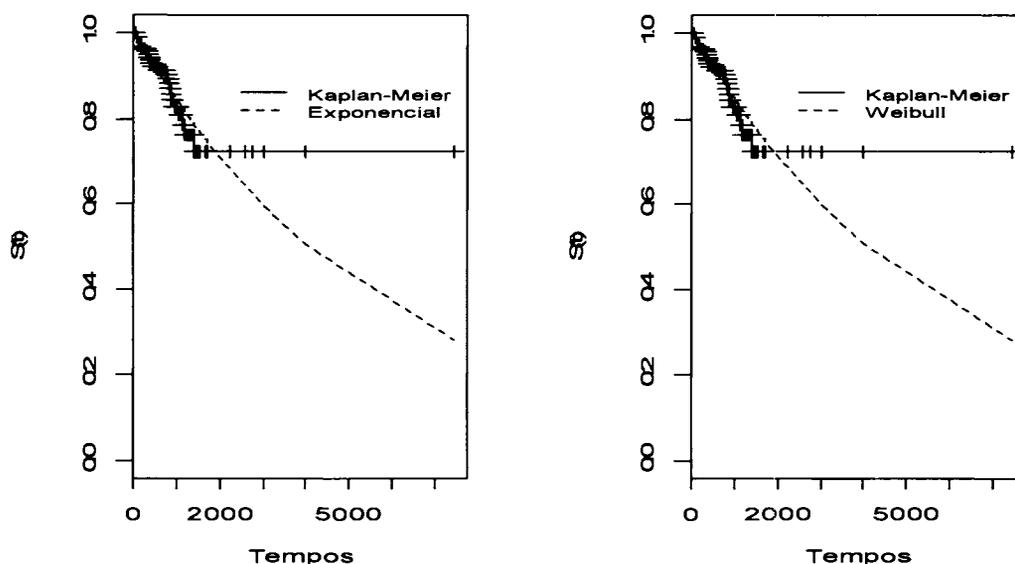


Figura 48. Curvas de sobrevivências estimadas pelos modelos de Exponencial e Weibull

Como observamos os modelos paramétricos não ajustam os dados então podemos concluir-se o modelo paramétrico não é adequado para estes dados. Existem várias situações para explicar porque o modelo paramétrico não se ajusta aos dados.

- a. O tamanho de amostra é pequeno
- b. Acontecem poucos eventos.

4.1.4. Modelos de Fragilidade

Neste trabalho, propomos uma generalização dos modelos de riscos proporcionais com interesse para ver se alguns heterogeneidade das covariáveis. O modelo de fragilidade dada por:

$$\lambda_i(t) = z_i \lambda_0(t) \exp \{ X_i' \beta \}$$

Consideramos os dados de sobrevivência univariados, para descrever a influência de covariáveis não observadas (heterogeneidade), em um conjunto de dados casamentos dos imigrantes. Consideramos o modelo com as variáveis que foram significativas na análise anterior: os motivos de entrada, União Europeia, nível escolaridade.

Observa-se que apenas o motivo de entrada 2 não apresenta efeito significativo, analogamente ao que se passou com o modelo de Cox. O modelo de fragilidade mostra-se não haver associação significativa entre as covariáveis, pois o valor de $p = 0,950$. Portanto as estimativas do modelo de Cox na análise anterior são na mesma e não se alterar os efeitos excepto ver na Tabela 30.

Tabela 31. Estimativas do modelo de fragilidade gama ajustados aos dados imigrantes

Covariável	Coefficiente	Erro Padrão	Valor p	IC 95%Exp β
Moterna 1	-1,163	0.567	0.0400	[0.1030; 0.9490 [
Moterna 2	0,762	0.596	0.2000	[0.6663; 6.8930 [
Nivescola 1	-1,485	0.490	0.0025	[0.0866; 0.5930 [
Nivescola 2	-1,699	0.731	0.02000	[0.0436; 0.7660 [
Pue	-1,350	0.571	0.0180	[0.0846; 0.7950 [
Fragilidade gama	-	-	0.9500	

Através do modelo fragilidade podemos concluir-se o modelo de fragilidade não difere do modelo de Cox, portanto o modelo de fragilidade não é significativo e mantém-se a conclusão do modelo Cox.

CAPÍTULO V

Conclusões e Recomendações

5.1. Conclusões

A imigração é um fluxo, com carácter permanente ou temporário, com a intenção de trabalho e/ou residência, de pessoas ou populações, de um país para outro. Trata-se de uma temática actual e com um interesse crescente dado o impacto que tem, quer nos países de origem, quer no país de acolhimento.

Neste trabalho realizámos um estudo sobre os imigrantes que entraram no distrito de Évora entre 2006 a 2009. Trata-se de 354 indivíduos, inscritos legalmente nos Serviços de Estrangeiros e Fronteiras de Évora. A maioria dos imigrantes é do sexo feminino, representando 68,4%. As nacionalidades são 39 sendo a mais representativa a brasileira, tendo também relevância alguns países da União Europeia (Holanda, Alemã e Espanha,) as nações africanas, em especial Cabo-Verde e Moçambique, dos países restantes destaca-se a Roménia e a Ucrânia. A maioria dos imigrantes vem de países que não fazem parte da União Europeia (63, 6%), Por outro lado podemos dizer que 348 Indivíduos mantiveram a nacionalidade de origem e 6 indivíduos adquiriram outra nacionalidade.

O nível de escolaridade mais representativo é o secundário (41,5%), sendo apenas um valor residual aqueles que não tem grau escolar (1,7%). A maioria dos imigrantes comunitários é oficialmente casada (65,1%), seguida pelos indivíduos solteiros (30,3%), e divorciados (2,9%). A maior concentração regista-se no concelho de Évora. Por outro lado, no distrito de Évora os indivíduos que residem em áreas urbanas são 54% e os imigrantes que residem nas áreas rurais são 46%.

Há vários motivos que trouxeram os imigrantes para Portugal sendo mais representativo o reagrupamento familiar com 53,3%, a maioria destes são reagrupamentos entre os imigrantes do sexo feminino com o seu cônjuge, e a proporção de reformados é muito baixa sendo de 1,6%. Os reformados, ao contrário dos restantes, são maioritariamente da União Europeia (França, Bélgica e Holanda). Os que têm emprego são 35,3%, e a maioria deles trabalham nas obras, e como empregadas em lares dos idosos. A minoria é constituída pelos imigrantes que estavam a procurar do 1º trabalho que representa 1,9%.

A nacionalidade dos cônjuges é na sua maioria Portuguesa, com uma percentagem de 76,7%, existindo ainda outras 14 nacionalidades, sendo a romena a segunda mais representativa com apenas 5,3%. Relativamente à ocupação dos cônjuges, 89,3% tem emprego, existindo apenas 1,1% de estudantes e 2,8% de reformados.

Um dos objectivos principais deste trabalho é analisar os casamentos dos imigrantes. A maior parte dos casais celebraram o casamento de tipo civil sem convenção antenupcial (45%). A seguir, o tipo de celebração do casamento civil (21,1%) e União de facto que representa 16,5%. A maior parte dos casais têm 1 filho.

Do estudo feito foi possível concluir que o género dos imigrantes depende dos países de origem (mais brasileiras e mais europeus do que seria de esperar se houvesse independência), sendo que há muito mais imigrantes do sexo feminino oriundos de fora da União Europeia do que haveria no caso de independência. Existe, pois, evidência estatística de que o género dos imigrantes depende dos países de origem. Concluímos que a maioria dos imigrantes do sexo feminino fora da União Europeia casou com os portugueses para melhorar a vida.

Estudando os motivos de entrada relativamente ao género obtém-se que o motivo de reagrupamento familiar é exibido em número superior pelo sexo feminino (127) em relação ao sexo masculino (39); podemos ainda dizer que o sexo feminino tem maior possibilidade ou maior número de motivos de entrada quando comparados com os motivos de entrada do sexo masculino. Podemos também concluir que os motivos de entrada dependem do ser oriundo ou não de um país da União Europeia.

Por outro lado, observou-se que o número de imigrantes que fazem parte da União Europeia inscritos no SEF de Évora cresceu entre 2006 e 2007, decrescendo entre 2008 e 2009. Tal ficou a dever-se ao facto desses imigrantes passarem a pedir os certificados na câmara municipal, e não passarem pelo SEF. Conhecidos os valores fornecidos pela Câmara verificámos que os números anteriores não permitiam avaliar correctamente o fluxo migratório. Na globalidade o que mostra a importância que tem o cruzamento de diferentes bases de dados. Os imigrantes que entraram mais neste quarto anos foram os Europeus, em seguida os brasileiros os restantes.

Comparando a idade dos imigrantes e idade dos seus cônjuges, conclui-se que a maioria dos imigrantes é casada com pessoas mais velhas.

Através da análise via Kaplan-Meier foi possível conhecer, identificando períodos maior risco, as variáveis que são importantes por si só para explicar o tempo de duração do

casamento dos imigrantes: celebração do casamento, oriundo de país da União Europeia, motivo de entrada, nível de escolaridade, nacionalidade e existência de filhos do casamento. Pode-se concluir com base nos dados que pudemos analisar através do modelo de regressão de Cox que os motivos de entrada, nível de escolaridade e país de origem (União Europeia) são variáveis explicativas da duração do casamento dos imigrantes.

Os imigrantes oriundos parte da União Europeia têm um risco 4 vezes superior de ruptura do casamento quando comparados com os imigrantes que não fazem parte da União Europeia. Por outro lado, para um indivíduo de nível de escolaridade básico o risco aumenta 4 vezes relativamente a um indivíduo com nível de escolaridade secundário. Para um indivíduo de nível escolaridade básico esse risco aumenta 6 vezes relativamente um indivíduo com nível superior de escolaridade. Portanto podemos concluir que um nível de escolaridade básico e o ser oriundo de um país da União Europeia são potenciadores do risco de cisão do casamento enquanto entrar por motivo de reagrupamento familiar é um factor protector.

Outra conclusão que retirei dessa análise é a de que os brasileiros e africanos se unem tendencialmente mais a parceiros portugueses (74% e 59% respectivamente), o que será eventualmente explicável pela procura de legalização no país através do casamento.

Na abordagem paramétrica tentámos os modelos Exponencial, Weibull, Log-normal e Log-logístico, mas nenhum dos modelos se mostrou adequado para esses dados. Também a abordagem introduzindo uma componente de fragilidade não conduziu a melhorias no modelo.

5.2. Recomendações

O SEF poderia solicitar mais elementos (inserir os dados no computador). O SEF devia tratar os dados adequadamente e atempadamente de modo a conhecer melhor a imigração em Portugal; esses estudos poderiam ser feitos a nível nacional e regional de modo a permitir a elaboração de estratégias de intervenção junto aos imigrantes.

Deveria melhorar o tratamento dos seus dados de modo a ter uma informação mais fidedigna e deve continuar a promover este tipo de estudos dando-lhe uma maior abrangência geográfica.

Nota-se que alguns imigrantes casam com portugueses só para legalizar o título da residência para viver em Portugal. Por isso, pode-se, recomendar:

1. O governo pode fazer uma lei mais dura para os imigrantes, a função do casamento. Por exemplo, alguns deles casam só para ganhar direito viver cá e depois disso deixou divorciam-se. Repare-se que é aos imigrantes oriundos de países da união europeia a quem pode interessar um casamento por conveniência⁴⁵.
2. Ao SEF para controlar os imigrantes que não tem motivos claros ou seja Deve ficar atento ao casamento dos imigrantes que entram por motivo de trabalho.

Por outro lado a aprendizagem da língua é fundamental neste processo de aculturação. A interpretação do que o outro refere como causa de algo que o está a incomodar, deverá ser o mais transparente possível para que se consiga iniciar a compreensão do fenómeno. Ou seja, se ao obstáculo da língua adicionarmos as crenças, os costumes, os valores e a cultura desse outro, teremos o nosso problema potencializado.

⁴⁵ <http://aeiou.expresso.pt/imigracao-sefnorte-detetou-25-a-30-casamentos-de-conveniencia-por-mes=f596521>.

BIBLIOGRAFIA

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*.
- Cardoso. F e Colosimo. E (2003), *Avaliação da suposição de riscos proporcionais para o modelo de Cox*, 48ª Reunião da RBRAS e 10º SEAGRO, Universidade Federal de Lavras, Brasil.
- Colosimo E. A, Giolo S.R. (2006), "*Análise Sobrevida Aplicada*", Edgard Blucher, Brasil.
- Dias Ana, (2004), "*Modelos de Tempo de Vida Acelerado em Análise Sobrevida*", tese, Faculdade ciências, Universidade de Lisboa.
- Falcão. Luísa (2002), "*A imigração em Portugal*" – Relatório síntese elaborado pelo Delta Consultores, Lisboa.
- Gageiro. João e Pestana. Maria (2009), "*Análise Categórica, Árvores de Decisão e Análise de Conteúdo – Em Ciências Sociais e da Saúde com o SPSS*", LIDEL – Edições Técnicas, Lda, Lisboa.
- H.Manuela e H. Andrew, (2009), *Investigação por Questionário* "Silabo, 2ª edição, Lisboa.
- Hosmer, Lemeshow, May (2008), *Applied Survival Analysis (Regression Modeling of Time-to-Event-Data*, 2nd Edition, New Jersey.
- Júnior Cláudio (2007), *Comportamento reprodutivo Uma análise a partir do grupo ocupacional das mulheres*, tese, Universidade Federal, Brasil.
- Machin, Cheung, Parmar (2006), *Survival Analysis A Practical Approach*, 2nd Edition, Enland, Jhon wiley & Sons, Ltd.
- Maroco, J. (2007), *Análise Estatística – Com Utilização do SPSS*, Edições Sílabo, Lda, 3ª edição, Lisboa.
- Malheiros J. Macaísta. (1966), "*Imigrantes na Região de Lisboa: os anos da Mudança*", Colibri, Lisboa.
- Meireles. J. M, Meireles. M.M, Barbosa. M.S, (2002), "*Guia Prático Jurídico Do Cidadão Imigrante (Não Comunitário): Direitos E Deveres Em Portugal*" 1ª edição, Associação Famílias, Braga.
- Michel. A, (1983), "*Sociologia da Família e do Casamento*", RÉ S, Coimbra.
- Oliveira A. Gouveia, (2009), "*Bioestatística Epidemiologia e Investigação*", lidel, Lisboa.
- O Impacto da Imigração nas sociedades Europa VI Port.pdf*. On line em <http://www.sef.pt/documentos/56/VersaoFinal>.

- Paulo, Raposo (2009), “ *Os Fluxos Matrimoniais Transnacionais entre Brasileiras e Portugueses Género e Imigração*”, Alto-Comissariado para a Imigração e Diálogo Intercultural (ACIDI, I. P.), Lisboa
- Pestana, Gageiro (2003), “*Análises de Dados para ciências Sociais; A Complementaridade do SPSS*”, Silabo, 3ª edição, Lisboa.
- Pestana, Gageiro (2005), *Análises de Dados para ciências Sociais; A Complementaridade do SPSS*, Silabo, 4ª edição, Lisboa.
- Rocha. G. P. Nunes, Medeiros. O. H. Ribeiro, Ferreira. E. (2009), *Perfis e Trajectórias dos Imigrantes nos Açores, Governo de Açores, Açores.*
- SEF, (2008), *Relatório de Imigração Fronteiras e Asilo*, Departamento e Planeamento de Formação.
- SOS Racismo (2002), *A imigração em Portugal: Os movimentos humanos e culturais em Portugal*. Lisboa: SOS Racismo.
- SOS Racismo (2005), *A imigração e Etnicidade: Vivências e Trajectórias de Mulheres em Portugal*. Lisboa: SOS Racismo.
- Turkman. M. A e Silva G.L. (2000), *Modelos Lineares Generalizados, Trabalho teoria e Prática, Universidade de Lisboa*, Lisboa.
- Trindade, (2001), *A Imigração*, Lisboa.
- <http://pt.wikipedia.org/wiki/Imigração>.
- SEF, *relatório anual de 2007*.
- <http://sosracismo.blogspot.com/>.
- <http://confrontos.no.sapo.pt/page9b.html>.
- Noticia Expresso sobre *o casamento por conveniência*. On line em <http://aeiou.expresso.pt/imigracao-sefnorte-detetou-25-a-30-casamentos-de-conveniencia-por-mes=f596521>.
- Revista de Estudos Demográficos, nº 43 – Junho 2008, on line em www.ine.pt.

ANEXOS

Anexo 1

Descrição das variáveis

❖ Nacionalidades dos imigrantes

Tabela 1. Distribuição por imigrante nacionalidade

Nacionalidade	Código
Brasileira	0
Cabo-verdiana	1
Ucraniana	2
Costa marfim	3
Boliviana	4
Guineense	5
Moçambicana	6
Francesa	7
Americana	8
Santomense	9
Portuguesa	10
Peruana	11
Romena	12
Colombiana	13
Moldova	14
Novazelândia	15
Angolana	16
Vietnamita	17
Chinesa	18
Filipinas	19
Holandesa	20

Tabela 2. Distribuição imigrante por nacionalidade continuação

Nacionalidade	Código
Venezuelana	21
Belga	22
Italiana	23
Alemã	24
Argentina	25
Lituana	26
Espanhola	27
Marroquina	28
Suíça	29
Búlgara	30
Russa	31
Rep. Checa	32
Finlandesa	33
Polaco	34
Irlandesa	35
Húngara	36
Britânica	37
Estoniana	38
Costa Rica	39

❖ **Naturalidade dos imigrantes e do Cônjuge**

Tabela 3. Distribuição imigrante por naturalidade

Naturalidade	Código
Brasil	0
Cabo-Verde	1
Ucrânia	2
Costa Marfim	3
Bolívia	4
Guine	5
Moçambique	6
França	7
América	8
São Tome	9
Portugal	10
Itália	11
Angola	12
Peru	13
Roménia	14
Colômbia	15
Moldova	16
Nova Zelândia	17
Vietname	18
China	19
Filipina	20
Holanda	21
Venezuela	22
Bélgica	23

Tabela 4. Distribuição imigrante por naturalidade continuação

Naturalidade	Código
Holanda	24
Argentina	25
Lituana	26
Espanha	27
Alemanha	28
Marrocos	29
Suíça	30
Bulgária	31
Rússia	32
Rep. Checa	33
Finlândia	34
Costa Rica	35
Hungria	36
Estónia	37
Reino Unido	38
Irlanda	39

❖ Estado Civil e Nível Escolaridade dos imigrantes

Tabela 5. Distribuição imigrante por estado civil

Estado Civil	Código
Solteira	0
Casado	1
Divorciada	2
Viúva	3
Separada	4
Outro	5

Tabela 6. Distribuição imigrante por nível escolaridade

Nível Escolaridade	Código
Nenhum	0
Básico	1
Secundário	2
Técnico Profissional	3
Superior	4

❖ Sexo e Países União Europeia

Tabela 8. Distribuição imigrante por países União Europeia

Países União Europeia	Código
Sim	0
Não	1

Tabela 7. Distribuição imigrante por sexo

Sexo	Código
Feminino	0
Masculino	1

❖ Motivo de Entrada e Profissão dos imigrantes

Tabela 9. Distribuição imigrante por Motivo de entrada

Motivo de Entrada	Código
Trabalho	0
Reagrupamento Familiar	1
Estudo	2
Reforma	3
Outro	4

Tabela 10. Distribuição imigrante por profissão

Profissão	Código
Empregado	0
A procura do 1º trabalho	1
Desempregado	2
Domestico	3
Reformado	4
Estudante	5
Outro	6

❖ **Celebração do Casamento, Distrito de residência e Designação da Morada**

Tabela 11. Distribuição imigrante por Celebração do casamento

Celebração do casamento	Código
Civil	0
Civil, sem convenção antenupcial	1
Civil, no regime adquiridos	2
União de Facto	3
Católico, sem convenção antenupcial	4
Civil, com convenção antenupcial	5
Civil, Sob regime da comunhão parcial	6

Tabela 12. Distribuição imigrante por Distrito

Distrito	Código
Évora	0
Beja	1
Portalegre	2

Tabela 13. Distribuição imigrante por designação da morada

Designação da Morada	Código
Rural	0
Urbano	1

❖ **Status/Eventos**

Tabela 14. Distribuição imigrante por status

Eventos	Código
Não Ocorreu Divorcio	0
Ocorreu Divorcio	1

❖ **Residência antes e depois de casamento**

Tabela 15. Distribuição imigrante por residência antes e depois de casamento

Residência	Código
Brasil	0
Cabo-Verde	1
Ucrânia	2
Costa Marfim	3
Bolívia	4
Guine	5
Moçambique	6
França	7
América	8
São Tome	9
Portugal	10
Reino Unido	11
Austria	12
Roméia	13
Colômbia	14
Moldova	15
Nova Zelândia	16
Angola	17
Vietname	18
China	19
Filipina	20
Itália	21

**Tabela 16. Distribuição por imigrante residência antes e depois de casamento
continuação**

Residência	Código
Holanda	22
Argentina	23
Lituana	24
Alemanha	25
Marrocos	26
Suíça	27
Rússia	28
Espanha	29
Bélgica	30
Rep.Checa	31
Bulgária	32
Finlândia	33
Irlanda	34
Costa Rica	35
Holanda	36
Escócia	37
Hungria	38
Estónia	39
Venezuela	40
Cuba	41

❖ **Residência antes e depois de casamento por classes**

Tabela 17. Distribuição imigrante por residência antes e depois de casamento por classes

Residência antes e depois casamento por classes	Código
País de Origem	0
País de Cônjuge	1
Outro País	2

As variáveis seguintes por já serem numéricas não foram codificadas

Início de Casamento = tempo do início do casamento ou entrar em Portugal

Fim do casamento = tempo do fim do casamento ou fim do estudo

Tempo = Duração do casamento

Idade dos imigrantes

Número dos filhos do casamento e número dos filhos do outro casamento

Anexo 2

Descrição das covariáveis para análise sobrevivência

Tabela 18. Descrição das covariáveis no estudo

Código	Descrição	Categorias
Sexo	Género dos imigrantes	0 se Feminino 1 se Masculino
Idadecat1 Idadecat2	Categoria das idades dos imigrantes	0 se < 33anos 1 se > 33anos 0 se < Q ₁ 1 se Q ₁ -Q ₂ 2 se Q ₂ -Q ₃ 3 se >Q ₃
Celebcasa	Tipo de celebração do casamento	0 se Civil 1 se Outros
Ec	Estado Civil	0 se Não são casados 1 se casados
Pue	União Europeia	0 se Sim 1 se Não
Moterna	Motivo de entrada	0 se Trabalho 1 se Reagrup. Familiar 2 se Outro
Nivescola	Nível Escolaridade	0 se Básico 1 se Secundário 2 se Superior
Desmor	Designação da Morada	0 se Rural 1 se Urbano
Profimg	Profissão dos imigrantes	0 se Empregado 1 se Estudante 2 se Outro
Nacimg	Nacionalidade dos imigrantes	0 se Brasileiras 1 se Africanas 2 se União Europeus 3 se Restantes países
Natcjpg	Naturalidade dos cônjuges	0 se Portugal 1 se Outros
Nficas	Existência de filhos do casamento	0 se não tem filhos 1 se tem filhos
NfotC	Existência de filhos de outro casamento	0 se não tem filhos 1 se tem filhos

Anexo 3

O Script do software R para Análise sobrevivência

```
dados<-read.table ("C:\\Users\\ZXC\\Documents\\casam.csv", sep=";", header=T)
attach (dados)
dados
```

##Análise Exploratório

```
summary(dados)
```

##Variáveis categoria

```
table(Antcod)
table(Sexo)
table(Nivescola)
table(Profimg)
table(Profcjg)
table(Pue)
table(Desmor)
table(Moterna)
table(Rac)
table(Rdc)
table(Sexo, Ant)
table(Celebcasa)
table(NfotC)
table(Nficas)
table(Cnacimg)
table(Distrito)
table(Natcjg)
table(Naccj)
table(Natimg)
table(IdimgC)
table(Nficascat)
```

##Codificar nº dos filhos

```
for (i in 1:165) {
```

```

if (Nficas[i]<=0) dados$Nfiscat[i]<-"0"
if (Nficas[i]>=1) dados$Nfiscat[i]<-"1"}
attach (dados)
dados
table(Nfiscat)

```

##Codificar nº dos filhos do outro casamento

```

for (i in 1:165) {
if (NfotC[i]<=0) dados$NfotCat[i]<-"0"
if (NfotC[i]>=1) dados$NfotCat[i]<-"1"}
attach (dados)
dados
table(NfotCat)

```

##Kaplan-Meier

```

require (survival)
km<-survfit(Surv(dados$Tempo, dados$Status))
Surv(dados$Tempo, dados$Status)
km_Antcod<-survfit(Surv(Tempo, Status)~Antcod)
km_Antcod
survdiff(Surv(Tempo, Status==1)~Antcod, rho=0)
survdiff(Surv(Tempo, Status==1)~Antcod, rho=1)
plot(km_Antcod, lty = 1:2, col = 1:2, ylab = "S(t)", xlab = "Período", conf.int = F)
legend(1, 0.4, c("2006 e 2007", "2008 e 2009"), lty = 1:2, col = 1:2)
title("Curvas de Sobrevida para Ano de entrada codificados")
Surv(dados$Tempo, dados$Status)
km_Nfiscat<-survfit(Surv(Tempo, Status)~Nfiscat)
km_Nfiscat
survdiff(Surv(Tempo, Status==1)~Nfiscat, rho=0)
survdiff(Surv(Tempo, Status==1)~Nfiscat, rho=1)
plot(km_Nfiscat, lty = 1:2, col = 1:2, ylab = "S(t)", xlab = "Período", conf.int = F)
legend(1, 0.3, c("Não tem Filhos", "Tem Filhos"), lty = 1:2, col = 1:2)
title("Curvas de Sobrevida a Números dos filhos do casamento")
Surv(dados$Tempo, dados$Status)
km_NfotCat<-survfit(Surv(Tempo, Status)~NfotCat)

```

```

km_NfotCat
survdif(Surv(Tempo, Status==1)~NfotCat, rho=0)
survdif(Surv(Tempo, Status==1)~NfotCat, rho=1)
plot(km_NfotCat, lty = 1:2, col = 1:2, ylab = "S(t)", xlab = "Período", conf.int = F)
legend(1, 0.3, c("Não tem Filhos", "Tem Filhos"), lty = 1:2, col = 1:2)
title("Curvas de Sobrevida Números filhos do outro casamento")
Surv(dados$Tempo, dados$Status)
km_IdimgC<-survfit(Surv(Tempo, Status)~IdimgC)
km_IdimgC
survdif(Surv(Tempo, Status==1)~IdimgC, rho=0)
survdif(Surv(Tempo, Status==1)~IdimgC, rho=1)
plot(km_IdimgC, lty = 1:3, col = 2:5, ylab = "S(t)", xlab = "Período", conf.int = F)
legend(100, 0.2, c("<=Q1", "Q1-Q2", " Q2-Q3", ">Q3"), lty = 1:3, col = 2:5, bty = "n")
title("Curvas de sobrevida para Idade dos imigrantes")

```

Aplicar o modelo de Cox para todas as variáveis

```

fit1<-coxph(Surv(Tempo, Status)~Sexo, data=dados, method="breslow", x=TRUE)
fit1
summary(fit1)
fit2<-coxph(Surv(Tempo, Status)~factor(Nivescola), data=dados, method="breslow",
x=TRUE)
fit2
summary(fit2)
fit2$loglik
fit3<-coxph(Surv(Tempo, Status)~factor(Profimg), data=dados, method="breslow",
x=TRUE)
fit3
summary(fit3)
fit4<-coxph(Surv(Tempo, Status)~Profcjg, data=dados, method="breslow", x=TRUE)
fit4
summary(fit4)
fit5<-coxph(Surv(Tempo, Status)~Pue, data=dados, method="breslow", x=TRUE)
fit5
summary(fit5)
fit5$loglik
fit6<-coxph(Surv(Tempo, Status)~Desmor, data=dados, method="breslow", x=TRUE)

```

```

fit6
summary(fit6)
fit8<-coxph(Surv(Tempo, Status)~Ec, data=dados, method="breslow", x=TRUE)
fit8
summary(fit8)
fit9<-coxph(Surv(Tempo, Status)~factor(Celebcasa), data=dados, method="breslow",
x=TRUE)
fit9
summary(fit9)
fit10<-coxph(Surv(Tempo, Status)~factor(Cnacing), data=dados, method="breslow",
x=TRUE)
fit10
summary(fit10)
fit11<-coxph(Surv(Tempo, Status)~factor(Natcjpg), data=dados, method="breslow",
x=TRUE)
fit11
summary(fit11)
fit12<-coxph(Surv(Tempo, Status)~Naccj, data=dados, method="breslow", x=TRUE)
fit12
summary(fit12)
fit13<-coxph(Surv(Tempo, Status)~Idimg, data=dados, method="breslow", x=TRUE)
fit13
summary(fit13)
fit14<-coxph(Surv(Tempo, Status)~Idcjpg, data=dados, method="breslow", x=TRUE)
fit14
summary(fit14)
fit15<-coxph(Surv(Tempo, Status)~factor(Moterna), data=dados, method="breslow",
x=TRUE)
fit15
summary(fit15)
fit15$loglik

```

##Codificar a Variável idade

```

for (i in 1:165) {
if(Idimg[i]<=19) dados$Idadecat2[i]<-"0"
if(Idimg[i]>19)& (Idimg[i]<=30) dados$Idadecat2[i]<-"1"

```

```

if(Idimg[i]>30)& (Idimg[i]<=36) dados$Idadecat2[i]<-"2"
If(Idimg[i]>36) dados$Idadecat[i]<-"3"}
attach (dados)
dados
for (i in 1:165) {
if (Idimg[i]<=33) dados$Idadecat[i]<-"0"
if (Idimg[i]>33) dados$Idadecat[i]<-"1"}
attach (dados)
dados
fit16<-coxph(Surv(Tempo, Status)~factor(Idadecat), data=dados, method="breslow",
x=TRUE)
fit16
summary(fit16)
fit17<-coxph(Surv(Tempo, Status)~factor(IdimgC), data=dados, method="breslow",
x=TRUE)
fit17
summary(fit17)
fit18<-coxph(Surv(Tempo, Status)~factor(Nficascat), data=dados, method="breslow",
x=TRUE)
fit18
summary(fit18)
fit19<-coxph(Surv(Tempo, Status)~factor(NfotCat), data=dados, method="breslow",
x=TRUE)
fit19
summary(fit19)
fit20<-coxph(Surv(Tempo, Status)~Antcod, data=dados, method="breslow", x=TRUE)
fit20
summary(fit20)

```

###Aplicar o Cox para as covariaveis que significativa

```

mod1<-coxph(Surv(Tempo,
Status)~factor(Moterna)+factor(Nivescola)+Pue+factor(Cnacing)+factor(Celebcasa)+factor
(Nficascat), data=dados, method="breslow", x=TRUE)
mod1
summary(mod1)
mod1$loglik

```

##Retirar a União Europeia

```
mod2<-coxph(Surv(Tempo,  
Status)~factor(Moterna)+factor(Nivescola)+factor(Cnacimg)+factor(Celebcasa)+factor(Nfic  
ascat), data=dados, method="breslow", x=TRUE)
```

```
mod2
```

```
summary(mod2)
```

```
mod2$loglik
```

##Tirar celebração do casamento

```
mod3<-coxph(Surv(Tempo,  
Status)~factor(Moterna)+factor(Nivescola)+factor(Cnacimg)+factor(Nficascat), data=dados,  
method="breslow", x=TRUE)
```

```
mod3
```

```
summary(mod3)
```

```
mod3$loglik
```

##Retirar a nacionalidade dos imigrantes

```
mod4<-coxph(Surv(Tempo, Status)~  
factor(Moterna)+factor(Nivescola)+factor(Nficascat), data=dados, method="breslow",  
x=TRUE)
```

```
mod4
```

```
summary(mod4)
```

```
mod4$loglik
```

##Entrar outra vez com todas as variáveis

```
mod5<-coxph(Surv(Tempo,  
Status)~factor(Moterna)+factor(Nivescola)+Pue+factor(IdimgC)+factor(Celebcasa)+factor(  
Cnacimg)+Sexo+Desmor+factor(Profimg)+factor(Profcjg), data=dados, method="breslow",  
x=TRUE)
```

```
mod5
```

```
summary(mod5)
```

```
mod5$loglik
```

##Considera como modelo Final

```
mod6<-coxph(Surv(Tempo, Status)~Ec+factor(Moterna)+factor(Nivescola)+Pue,  
data=dados, method="breslow", x=TRUE)
```

```
mod6
```

```
summary(mod6)
```

```
mod6$loglik
```

##Modelo com interacção não é significativa

```
mod7<-coxph(Surv(Tempo,  
Status)~factor(Moterna)+factor(Moterna)*Ec+factor(Nivescola)+Pue+factor(Nivescola)*Pu  
e, data=dados, method="breslow", x=TRUE)  
mod7  
summary(mod7)  
mod7$loglik
```

Os resíduos Deviance e Martingal

```
cox.zph(mod5, transform="identity")  
par (mfrow=c(2,3))  
plot (cox.zph (mod5))  
rd<-resid(mod5,type="deviance") # resíduos deviance  
rm<-resid(mod5,type="martingale") # resíduos martingal  
p1<-mod5$linear.predictors  
plot(p1,rm, xlab="Preditor linear", ylab="Resíduo martingal", pch=16)  
plot(p1,rd, xlab="Preditor linear", ylab="Resíduo deviance" , pch=16)  
res.sch <- cox.zph(mod5)  
res.sch  
par(mfrow=c(2,3))  
plot(res.sch[1])  
abline(h = mod5$coef[1], lty = 2)  
plot(res.sch[2])  
abline(h = mod5$coef[2], lty = 2)  
plot(res.sch[3])  
abline(h = mod5$coef[3], lty = 2)  
plot(res.sch[4])  
abline(h = mod5$coef[4], lty = 2)  
plot(res.sch[5])  
abline(h = mod5$coef[5], lty = 2)  
plot(res.sch[6])  
abline(h = mod5$coef[6], lty = 2)
```

###resíduos padronizados de Schoenfeld

```
resid(mod5,type="scaledsch")
rsc<-cox.zph(mod5)    ### g(t) = t
rsc
par(mfrow=c(2,4))
plot(cox.zph(mod5))
plot(rsc[2])
abline(h=fit1$coef[2], lty=3)
Ht<-basehaz(mod5,centered=F)
tempos<-Ht$Tempo
H0<-Ht$hazard
S0<- exp(-H0)
round(cbind(tempos, S0,H0),digits=5)
par(mfrow=c(1,2))
rd<-resid(mod5,type="deviance")    # resíduos deviance
rm<-resid(mod5,type="martingale")  # resíduos martingala
p1<-mod5$linear.predictors
plot(p1,rm, xlab="Preditor linear", ylab="Resíduo martingal", pch=16)
plot(rm)
abline(h=0, col=2, lty=2)
plot(p1,rd, xlab="Preditor linear", ylab="Resíduo deviance" , pch=16)
plot(mod5, rm, xlab = "Estado Civil", ylab = "Resíduo")
lines(lowess(Ec, rm, iter = 0), lty = 2)
plot(p1,rd, xlab="Preditor linear", ylab="Resíduo deviance" , pch=16)
par(mfrow=c(2,3))
dfbetas<-resid(mod5,type="dfbeta") #resíduos score
plot(Ec,dfbetas[,1], xlab="Ec", ylab="Influência para Ec")
plot(factor(Moterna), dfbetas[,2], xlab="Moterna", ylab="Influência para Moterna")
plot(factor(Pue), dfbetas[,3], xlab="Pue", ylab="Influência para Pue")
plot(factor(Nivescola),dfbetas[,4], xlab="Nivescola", ylab="Influência para Nivescola")
```

###Pontos Influentes

```
infl<-apply(dfbetas,1,sum)
plot(1:mod5$n,infl, xlab="Índice",ylab="Influência")
abline(h=0)
```

```
a<-identify(infl)
```

##Os residuos

```
plot(1:mod5$n, dfbetas[, 5], xlab="Índice", ylab="Resíduos")
```

```
plot(1:mod5$n, dfbetas[, 4], xlab="Índice", ylab="Resíduos")
```

```
a<-identify(dfbetas)
```

##Análises dos modelos Modelos Parâmetros

```
dados<-read.table("C:\\Users\\ZXC\\Documents\\casam.csv", sep=";", header=T)
```

```
attach(dados)
```

```
dados
```

```
require(survival)
```

#linearização para Exp, Weib e Log-normal

```
i<-order(dados$Tempo)
```

```
ekm<- survfit(Surv(dados$Tempo,dados$Status)~1)
```

```
summary(ekm)
```

```
st<-ekm$surv
```

```
temp<-ekm$time
```

```
invst<-qnorm(st)
```

```
par(mfrow=c(1,3))
```

```
plot(temp, -log(st),pch=16,xlab="Tempos",ylab="-log(S(t))")
```

```
plot(log(temp),log(-log(st)),pch=16,xlab="log(tempos)",ylab="log(-log(S(t))")
```

```
plot(log(temp),invst,pch=16,xlab="log(tempos)",ylab=expression(Phi^-1*(S(t))))
```

estimativas de km vs estimativas dos modelos paramétricos

```
ajust1<-survreg(Surv(Tempo,Status)~1,dist='exponential')
```

```
ajust1
```

```
alpha<-exp(ajust1$coefficients[1])
```

```
alpha
```

```
ajust2<-survreg(Surv(Tempo,Status)~1,dist='weibull')
```

```
ajust2
```

```
alpha<-exp(ajust2$coefficients[1])
```

```
gama<-1/ajust2$scale
```

```
cbind(gama, alpha)
```

```
ajust3<-survreg(Surv(Tempo,Status)~1,dist='lognorm')
ajust3
ajust4<-survreg(Surv(Tempo,Status)~1,dist='loglogistic')
ajust4
```

##Análise Deviance

```
anova(ajust1, ajust2, ajust3, ajust4, test="Chisq")
anova(ajust3, ajust4, test="Chisq")
ajust1$loglik[2]
ajust2$loglik[2]
ajust3$loglik[2]
ajust4$loglik[2]
ekm<-survfit(Surv(Tempo,Status)~1)
time<-ekm$time
st<-ekm$surv
ste<- exp(-time/alpha)
stw<- exp(-(time/alpha)^gama)
stln<- pnorm((-log(time)+ 8.656)/ 1.774)
cbind(time,st,ste,stw,stln)
par(mfrow=c(1,4))
plot(st,ste,pch=16,ylim=range(c(0.0,1)), xlim=range(c(0,1)), xlab = "S(t): Kaplan-Meier",
      ylab="S(t): exponencial")
lines(c(0,1), c(0,1), type="l", lty=1)
plot(st,stw,pch=16,ylim=range(c(0.0,1)), xlim=range(c(0,1)), xlab = "S(t): Kaplan-Meier",
      ylab="S(t): Weibull")
lines(c(0,1), c(0,1), type="l", lty=1)
plot(st,stln,pch=16,ylim=range(c(0.0,1)), xlim=range(c(0,1)), xlab = "S(t): Kaplan-Meier",
      ylab="S(t): log-normal")
lines(c(0,1), c(0,1), type="l", lty=1)
plot(st,stln,pch=16,ylim=range(c(0.0,1)), xlim=range(c(0,1)), xlab = "S(t): Kaplan-Meier",
      ylab="S(t): loglogistic")
lines(c(0,1), c(0,1), type="l", lty=1)
# #Comparação das curvas de sobrevivência dos modelos com as curvas de
sobrevivência de KM
par(mfrow=c(1,2))
```

```

plot(ekm, conf.int=F, xlab="Tempos", ylab="S(t)")
lines(c(0,time),c(1,ste), lty=2)
legend(2000,0.9,lty=c(1,2),c("Kaplan-Meier", "Exponencial"),bty="n",cex=0.8)
plot(ekm, conf.int=F, xlab="Tempos", ylab="S(t)")
lines(c(0,time),c(1,stw), lty=2)
legend(2000,0.9,lty=c(1,2),c("Kaplan-Meier", "Weibull"),bty="n",cex=0.8)
plot(ekm, conf.int=F, xlab="Tempos", ylab="S(t)")
lines(c(0,time),c(1,stln), lty=2)
legend(2000,0.9,lty=c(1,2),c("Kaplan-Meier", "Log-normal"),bty="n",cex=0.8)
plot(ekm, conf.int=F, xlab="Tempos", ylab="S(t)")
lines(c(0,time),c(1,stln), lty=2)
legend(2000,0.9,lty=c(1,2),c("Kaplan-Meier", "Loglogistic"),bty="n",cex=0.8)

```

Ajustamento do Modelo Weibull

```

Ec<-factor(Ec)
Moterna<-factor(Moterna)
Nivescola<-factor(Nivescola)
Pue<-factor(Pue)
ajustln<-survreg(Surv(Tempo,Status) ~ Ec+Moterna+Nivescola+Pue, dist="weibull")
summary(ajustln)
ajustln$loglik

```

Teste AIC para compara o modelo

```

extractAIC (ajust1)
extractAIC (ajust2)
extractAIC (ajust3)
extractAIC (ajust4)

```

Os Resíduos

```

# resíduos de Cox-Snell
i<-order(dados$Tempo)
t<-dados$Tempo
res<- t*exp(-ajustwb$linear.predictors)
ekm <- survfit(Surv(res,dados$Status)~1,type=c("kaplan-meier"))
summary(ekm)

```

```

par(mfrow=c(1,2))
plot(ekm, conf.int=F,lty=c(1,1),xlab="resíduos",ylab="S(e) estimada")
res<-sort(res)
exp<-exp(-res)
lines(res,exp,lty=3)
legend(2,0.8,lty=c(1,3),c("Kaplan-Meier", "Exponencial(1)"),lwd=1,bty="n",cex=0.7)
st<-ekm$surv
t<-ekm$time
sexp<-exp(-t)
plot(st,sexp,xlab="S(e) - Kaplan-Meier", ylab= "S(e) - Exponencial(1)",pch=16)

```

##Resíduos Deviance e Martingal

```

par(mfrow=c(1,2))
rd<-resid(ajustwb,type="deviance")    # resíduos deviance
rm<-resid(ajustwb,type="martingale")  # resíduos martingala
p1<-ajustwb$linear.predictors
plot(ajustwb,rm, xlab="Preditor linear", ylab="Resíduo martingal", pch=16)
plot(rm)
abline(h=0, col=2, lty=2)
plot(ajustwb,rd, xlab="Preditor linear", ylab="Resíduo deviance" , pch=16)

```

Resíduos Padronizados

```

xb<-
ajust2$coefficients[1]+ajust2$coefficients[2]*Ec+ajust2$coefficients[3]*Moterna+ajust2$coefficients[4]*Nivescola+ajust2$coefficients[5]*Pue
sigma<-ajust2$scale
res<-((log(Tempo)-(xb))/sigma)
resid<-exp(res)
ekm<-survfit(Surv(resid,Status)~1)
resid<-ekm$Tempo
sln<-pnorm(-log(resid))
par(mfrow=c(1,2))
plot(ekm$surv,sln,xlab="S(ei*):Kaplan Meier",ylab="Weibull",pch=16)
plot(ekm,conf.int=F,mark.time=F,xlab="Resíduos(ei*)",ylab="Sobrevivência Estimada",pch=16)
lines(resid,sln,lty=2)

```

```

modeloweib <- survreg(Surv(Tempo, Status) ~ 1, data = dados, dist = "weib")
summary(modeloweib)
modeloexp <- survreg(Surv(Tempo, Status) ~ 1, data = dados, dist = "exp")
summary(modeloexp)
modlm<-survreg(Surv(Tempo,Status)~1,dist='lognorm') #ajustamento do modelo
summary(modlm)
modloglogistic<-survreg(Surv(Tempo,Status)~1, dist="loglogistic")
summary(modloglogistic)
## Modelos Fragilidade
dados<-read.table ("C:\\Users\\ZXC\\Documents\\casam.csv", sep=";", header=T)
attach (dados)
dados
require (survival)
id<-1:165
km<-survfit(Surv(dados$Tempo, dados$Status))
mod5<-coxph(Surv(Tempo,
Status)~Ec+factor(Moterna)+factor(Nivescola)+Pue+frailty(id,dist="gamma"), data=dados,
method="breslow", x=TRUE)
mod5
summary(mod5)
wi<-mod5$frail
zi<-exp(wi)
plot(id, zi, xlab = "Imigrantes (1 a 165)", ylab="zi estimados", pch=16)
abline(h=1,lty=2)
ne(h=1,lty=2)

```