



## UNIVERSIDADE DE ÉVORA

### ESCOLA DE CIÊNCIAS E TECNOLOGIA

#### DEPARTAMENTO DE MATEMÁTICA

**Modelação estatística do tempo de duração de procedimentos nos serviços académicos de uma Universidade**

**Rita Brazão de Freitas**

Orientação: Professor Doutor Paulo Jesus Infante dos Santos, Universidade de Évora  
Coorientação: Professor Doutor Gonçalo João Costa Jacinto, Universidade de Évora

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

Área de especialização: *Modelação Estatística e Análise de Dados*  
Dissertação

Évora, 2015

## **Modelação estatística do tempo de duração de procedimentos nos Serviços Académicos de uma Universidade**

Os Serviços Académicos são fundamentais para o bom funcionamento da Universidade de Évora e, consequentemente, para a imagem que esta projeta para o exterior. Nesta dissertação, pretendeu-se estudar os tempos de duração de alguns procedimentos iniciados pelos alunos nestes serviços. No caso particular dos Mestrados e Doutoramentos, a entrega do projeto de dissertação e todos os procedimentos que se seguem até à realização das provas são fulcrais para a conclusão do curso no tempo previsto. Para modelar o tempo até à entrega da dissertação, recorremos à análise de sobrevivência. Com base num modelo de regressão logística, procurámos traçar o perfil mais provável do aluno que necessita de pedir semestres adicionais para terminar a sua dissertação de Mestrado. Para a análise dos restantes procedimentos iniciados pelos alunos, recorremos ao controlo estatístico de qualidade, discutindo e elaborando algumas cartas de controlo e calculando índices de capacidade do serviço.

## **Statistical modeling of the procedures duration in a University's Academic Services**

The Academic Services are crucial to the proper functioning of Évora University and, consequently, to the image that it projects outwards. In this thesis we intended to study the duration of some proceedings initiated by students in Academic Services. In the particular case of the Master's and Doctor's Degree, the delivery of the thesis project and all the procedures that follows until the final examination are critical to complete the degree on time. To model the time until the delivery of the final thesis, we resorted to survival analysis and, based on a logistic regression model, we tried to draw the most likely profile of the student who needs to ask for additional semesters to finish his thesis. For the analysis of the remaining procedures we resorted to statistical quality control, discussing and constructing charts of retrospective quality control and measuring rates of service capacity.

## **Agradecimentos**

Ao Professor Doutor Paulo Infante, orientador desta tese de Mestrado, por todo o apoio e disponibilidade demonstrados, não só ao longo da elaboração desta tese, mas também no decorrer de todo o Mestrado. A sua contribuição foi fundamental para a minha formação académica. Agradeço também a oportunidade que me proporcionou de integrar outros projetos e reconheço, com gratidão, a confiança que depositou em mim desde o início deste Mestrado.

Ao Professor Doutor Gonçalo Jacinto, coorientador desta tese de Mestrado, pelo seu constante incentivo, disponibilidade e apoio demonstrados. Agradeço também por tudo o que me ensinou ao longo do curso de Mestrado e ao longo da elaboração desta tese.

À Professora Doutora Fernanda Figueiredo (Faculdade de Economia da Universidade do Porto), pelo importante contributo no desenvolvimento desta tese, nomeadamente na área de Controlo Estatístico de Qualidade.

Ao Doutor João Dias (Serviços Académicos da Universidade de Évora), pela sua disponibilidade e paciência em atender todas as dúvidas que surgiram ao longo do tratamento dos dados e ao longo da elaboração desta tese.

À Vice Reitora para o Ensino e Formação, Professora Hermínia Vilar, e à Sra. Diretora dos Serviços Académicos, Dra. Margarida Cabral, cargos que desempenhavam na altura da entrega do projeto da dissertação, pelo interesse demonstrado nesta dissertação, permitindo a recolha e o tratamento dos dados. Agradeço também aos Serviços de Informática pela recolha e envio das bases de dados e, em particular, à Dra. Sónia Costa pela eficiência e profissionalismo demonstrados ao longo do processo de construção e interpretação da base de dados, bem como ao Engenheiro Luís Quintano pela celeridade de resposta e profissionalismo mostrados no envio dos dados usados na fase final da dissertação.

Aos restantes professores do Mestrado em Modelação Estatística e Análise de Dados, em especial à Vice Reitora para a Educação, Formação Graduada e Pós-graduada, Professora Doutora Maria Filomena Mendes, que muito contribuiu para a minha formação académica.

Aos meus colegas de Mestrado, pelos momentos partilhados.

---

## Agradecimentos

---

À minha família, em especial aos meus pais, por acreditarem em mim e nas minhas escolhas. Agradeço a vossa dedicação e todas as oportunidades que me proporcionaram ao longo dos anos. Espero que esta etapa possa, de alguma forma, retribuir e compensar todo o apoio demonstrado.

## Índice

Capítulo I – Introdução Geral .....	1
I.1. Contextualização .....	1
I.2. Objetivos gerais.....	2
Capítulo II – Material e Métodos.....	5
II.1. Caracterização da amostra.....	5
II.2. Metodologia aplicada.....	7
II.3. Pesquisa Bibliográfica.....	9
Capítulo III – Modelação do tempo até à conclusão de uma dissertação .....	15
III.1. Introdução.....	15
III.2. Metodologia aplicada.....	15
III.3. Caracterização da amostra.....	22
III.3.1. Cursos de Mestrado.....	22
III.3.2. Cursos de Doutoramento .....	27
III.4. Análise de Sobrevida – Mestrados.....	30
III.4.1. Estimativas de Kaplan-Meier .....	30
III.4.2. Modelo de Cox.....	35
III.4.3. Resultados .....	42
III.5. Análise de Sobrevida – Doutoramentos .....	45
III.5.1. Estimativas de Kaplan-Meier .....	45
III.5.2. Resultados .....	48
III.6. Conclusões .....	48
Capítulo IV – Perfil dos alunos que pedem semestres adicionais .....	51
IV.1. Introdução .....	51
IV.2. Metodologia aplicada .....	51
IV.3. Caracterização da amostra .....	56
IV.4. Modelo de Regressão Logística.....	61
IV.5. Resultados.....	67
IV.6. Conclusões .....	72
Capítulo V – Controlo de Qualidade dos tempos de procedimentos iniciados pelos alunos .....	73
V.1. Introdução .....	73
V.2. Metodologia aplicada .....	74
V.2.1. Cartas de Controlo Shewhart .....	75
V.2.2. Cartas de Controlo EWMA.....	77
V.2.3. Cartas de Controlo <i>Skew-Normal</i> .....	78
V.2.4. Análise de capacidade do processo.....	80
V.3. Caracterização da amostra .....	82
V.4. Análise dos dados .....	91
V.4.1. Tempo entre a entrega e discussão da dissertação .....	91
V.4.2. Tempo entre a entrega e o envio da dissertação para a escola .....	101
V.4.3. Tempo de formação do júri.....	104
V.4.4. Tempo até à primeira reunião do júri .....	107
V.4.5. Tempo entre a primeira reunião e a realização das provas .....	110
V.4.6. Tempo de espera pela emissão de diplomas .....	111

V.5. Conclusões .....	115
Capítulo VI – Considerações finais.....	117
Capítulo VII – Referências.....	121
ANEXO 1. Glossário.....	123
ANEXO 2. Tabelas .....	124

## **Capítulo I – Introdução Geral**

### **I.1. Contextualização**

Nos últimos anos, algumas das Universidades Europeias têm vindo a sofrer alterações orçamentais. Segundo o relatório apresentado pela associação Europeum (2012), Portugal tem vindo a acompanhar as tendências negativas de alguns países da Europa, nomeadamente a Espanha, Itália e a República Checa. Para além dos cortes aplicados recentemente, as Universidades Portuguesas deparam-se com outros problemas. Aliado ao facto de nos últimos anos assistirmos a um decréscimo do número de alunos com ingresso nas Universidades Portuguesas ou a um aumento do número de alunos que não terminam o seu curso superior, está o prolongamento da passagem dos alunos pelas Universidades. Qualquer uma das situações referidas pode ter um grande impacto, não só no orçamento das Universidades, mas também no orçamento das famílias que suportam os estudantes no decorrer do seu percurso académico. O futuro orçamento das Universidades será também função do número de diplomados em cada ano, pelo que um adiamento da conclusão do curso pode ter um impacto bastante significativo na sobrevivência da instituição. Segundo Quinn (2013), as Universidade Europeias devem, para além de proporcionar ações que possam melhorar a sua eficiência, implementar políticas de intervenção para prevenir os estudantes com comportamentos desfavoráveis no ensino superior.

Para que a Universidade de Évora possa acompanhar as Universidades de países mais bem sucedidos nesta matéria, como a Dinamarca, o Reino Unido e a Alemanha (Quinn, 2013), primeiro é necessário perceber quais as dinâmicas dos fluxos escolares dos alunos e quais os comportamentos dos serviços da própria Universidade. Neste sentido, a informação gerada pelos Serviços Académicos da Universidade de Évora (SAC) é extremamente importante na procura de uma melhoria contínua de modo a que a Universidade possa atingir níveis de excelência.

Os SAC constituem uma unidade de serviços cujo principal objetivo é a organização e gestão de assuntos de natureza académica. Estes serviços exercem a sua atividade nos domínios da informação, da organização e acompanhamento dos processos escolares dos alunos. Efetivamente, são os SAC que prestam informações sobre todos os assuntos relacionados com matrículas, inscrições, frequências e aproveitamentos dos alunos. Também está a seu cargo a organização dos processos dos candidatos a ingressos e reingressos, transferências, mudanças de cursos e equivalências. Para além disso, estes serviços são responsáveis pela organização dos processos administrativos de mestrados, doutoramentos e outros cursos de formação avançada e dos processos respeitantes às provas para obtenção do título de agregado. É ainda da sua competência

emitir certificados de habilitações e declarações e assegurar o atendimento aos alunos, docentes e ao público em geral. Trata-se, pois, de um serviço fundamental para o bom funcionamento da Universidade de Évora, sendo extremamente importante no apoio aos diversos serviços da Universidade, à Reitoria e ao seu gabinete de promoção de qualidade.

Os diferentes procedimentos iniciados pelos alunos nos SAC, estejam estes relacionados com a entrega de uma dissertação, com um pedido de semestre adicional ou com a emissão de um certificado, podem ter um grande impacto no orçamento da Universidade e na imagem que esta projeta para o exterior através dos seus alunos. No caso particular dos Mestrados e Doutoramentos, uma variável fulcral que pode adiar a sua conclusão no tempo previsto é o tempo de duração de uma dissertação.

Após a entrega do projeto de dissertação nos SAC, é desencadeado todo um processo que termina no momento em que os alunos realizam as provas públicas. Porém, este processo não depende única e exclusivamente dos alunos, mas também de um conjunto de atores que, de uma forma ou outra, podem antecipar ou adiar o tempo no qual todo o processo é concluído. É, pois, importante tentar perceber quais os fatores que estão na origem do adiamento da conclusão de alguns cursos. Será que este adiamento está relacionado com características sociodemográficas ou poderá o *background* familiar dos alunos ser mais importante? Que efeito poderá ter o percurso escolar anterior dos alunos? Será que alguns cursos desta Universidade são mais propensos a este adiamento? Ou, por outro lado, estará este adiamento relacionado com os processos da própria Universidade, tais como os tempos de duração da formação ou reunião do júri? Tomando como base a informação gerada pelos SAC e outros serviços da Universidade de Évora, aplicamos várias metodologias estatísticas na análise e modelação dos tempos de duração de alguns procedimentos iniciados pelos alunos.

## I.2. Objetivos gerais

Nesta dissertação, focámo-nos essencialmente nos processos que se iniciam a partir do momento em que os alunos entregam o seu projeto de dissertação nos SAC da Universidade. Começámos por estudar os tempos que decorrem entre a data da aprovação do projeto e a data da entrega da dissertação (de um curso de Mestrado ou de Doutoramento). Porém, nem todos os alunos entregam as suas dissertações no tempo previsto. Alguns precisam pedir pelo menos um semestre adicional para poder entregar a dissertação. Com vista nesta problemática, primeiro optámos por recorrer à análise de sobrevivência para modelar o tempo desde o registo da

dissertação até à data da sua entrega, considerando como possíveis variáveis explicativas algumas variáveis sociodemográficas e de natureza académica. Depois, para tentarmos encontrar determinantes que levam alguns alunos a pedir semestres adicionais, tentámos averiguar qual o perfil mais provável do aluno que efetua este pedido, recorrendo a um modelo de regressão logística. Neste último caso, considerámos apenas os alunos de Mestrado, pois o número de registo de alunos de Doutoramento era muito pequeno para realizar este tipo de abordagem.

Após a entrega da dissertação, outros procedimentos administrativos podem atrasar ainda mais o tempo até à conclusão do Mestrado. Estes procedimentos estão relacionados com os próprios SAC, com a comissão do curso ou até com os docentes dos diferentes cursos. O regulamento interno da Universidade estabelece prazos para cada um dos procedimentos administrativos. Contudo, sabemos que, para além das variações naturais a estes processos, existem causas responsáveis por uma melhor ou pior prestação dos mesmos. Neste contexto, recorremos ao controlo estatístico de qualidade, elaborando de uma forma crítica cartas de controlo (retrospectivas) e calculando índices de capacidade do serviço. Com esta metodologia analisamos os seguintes tempos:

1. O tempo entre a entrega da dissertação e a realização de provas públicas (T proc);
2. O tempo entre a entrega da dissertação e o seu envio para a escola (T sac);
3. O tempo de formação do júri (T formação júri);
4. O tempo desde a formação do júri até à realização da primeira reunião (T reunião);
5. O tempo desde a primeira reunião até à realização de provas públicas (T provas);
6. O tempo de espera pela emissão de um diploma de grau (TD grau) e diploma de curso (TD curso).

A implementação de cartas de controlo de qualidade, além de possibilitar a avaliação e controlo do desempenho dos diferentes processos ao longo do tempo, permite a distinção entre a variação natural do processo e outras formas de variação. Recorremos às cartas de controlo para monitorizar variáveis relativas aos procedimentos administrativos, pois também estas têm um grande impacto na imagem que o aluno projeta da Universidade para o exterior.

Na Figura 1.1, apresentamos a estrutura das análises efetuadas nesta dissertação, assim como a ordem cronológica dos procedimentos iniciados pelos alunos nos SAC da Universidade de Évora.

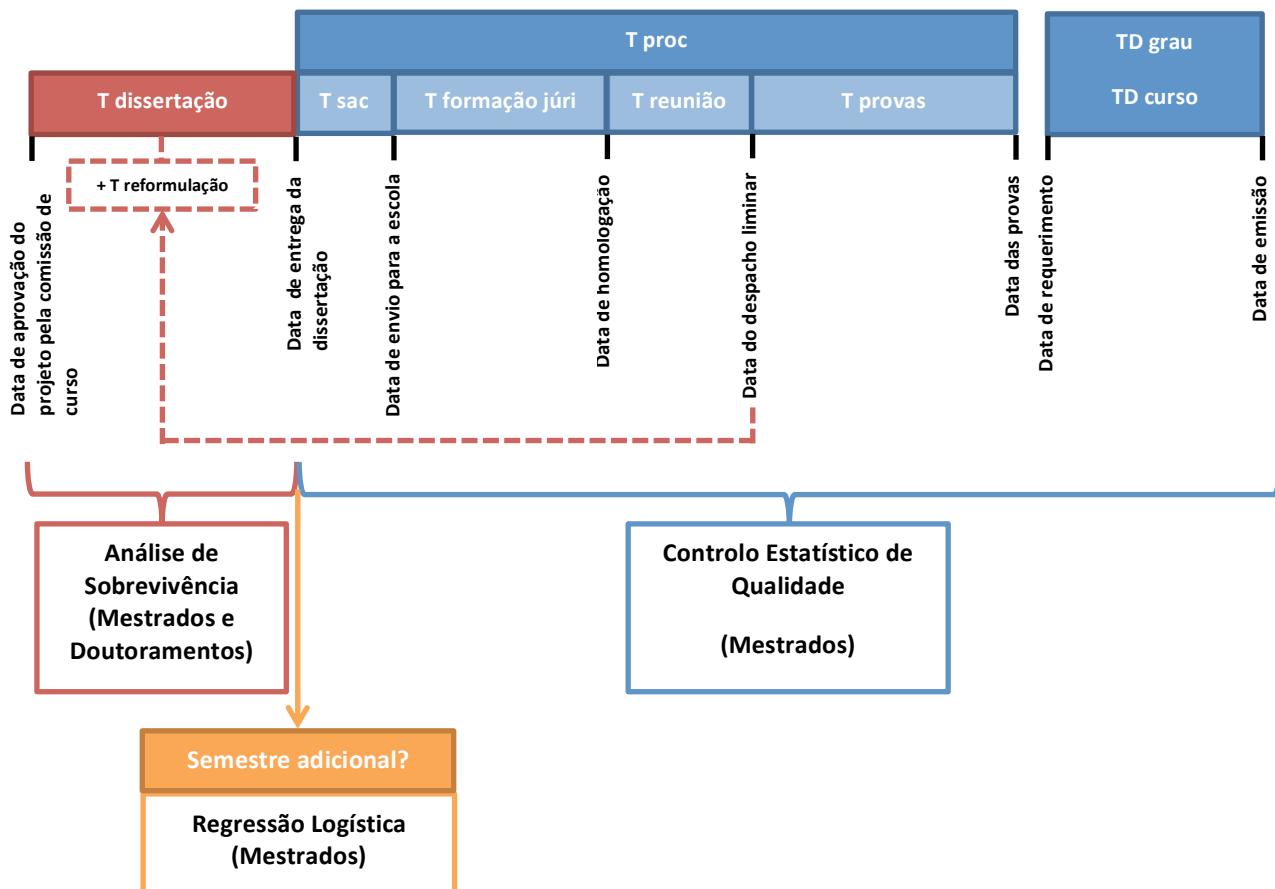


Figura 1.1: Estrutura das análises efetuadas, por ordem cronológica dos procedimentos iniciados pelos alunos nos SAC.

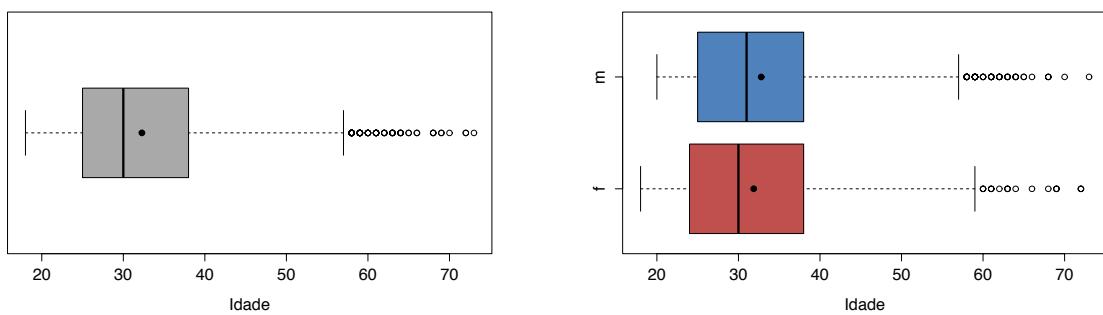
No segundo capítulo desta dissertação fazemos uma breve descrição da amostra e dos métodos utilizados. No terceiro capítulo, recorremos à análise de sobrevida para modelar o tempo desde o registo da dissertação até à data da sua entrega. No capítulo seguinte, procuramos os determinantes que levam alguns alunos a pedir semestres adicionais para terminar um curso de Mestrado, através de um modelo de regressão logística. Finalmente, no quinto capítulo, recorremos às cartas de controlo para monitorizar variáveis relativas aos procedimentos administrativos e medimos índices de capacidade dos serviços.

## Capítulo II – Material e Métodos

### II.1. Caracterização da amostra

A recolha dos dados necessários para esta dissertação foi efetuada pelos Serviços de Informática da Universidade de Évora. A amostra é constituída pelos alunos de Mestrados e Doutoramentos que ingressaram na Universidade de Évora entre os anos letivos de 2007/2008 e 2012/2013, tendo sido seguidos até Julho de 2013.

Dos 7116 alunos que ingressaram na Universidade no período de tempo referido, a maioria é do sexo feminino (56,4%) e de nacionalidade Portuguesa (77,2%). Estes alunos ingressaram na Universidade entre os 18 e 73 anos, sendo que 50% dos alunos fizeram-no até aos 30 anos (Figura 2.1). A média das idades destes alunos é igual a 32,3 anos e o desvio padrão igual a 9,2 anos. Em média, os homens são cerca de um ano mais velhos do que as mulheres. A idade média dos homens é igual a 32,8 anos (desvio padrão igual a 9,3 anos) e a idade média das mulheres é igual a 31,9 anos (desvio padrão igual a 9,1 anos), sendo que metade dos homens tem até 31 anos e metade das mulheres tem até 30 anos.

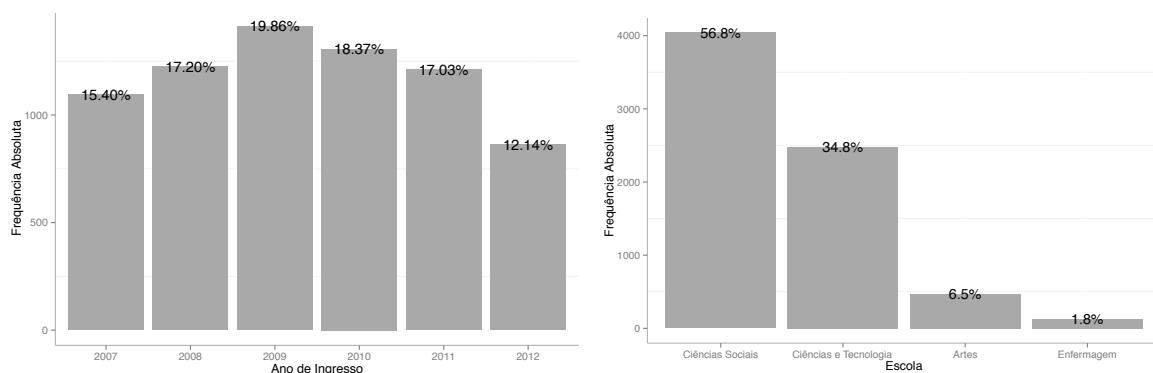


**Figura 2.1: Distribuição das idades dos alunos com ingresso em cursos de Mestrado ou Doutoramento no período de estudo; Distribuição das idades por sexo.**

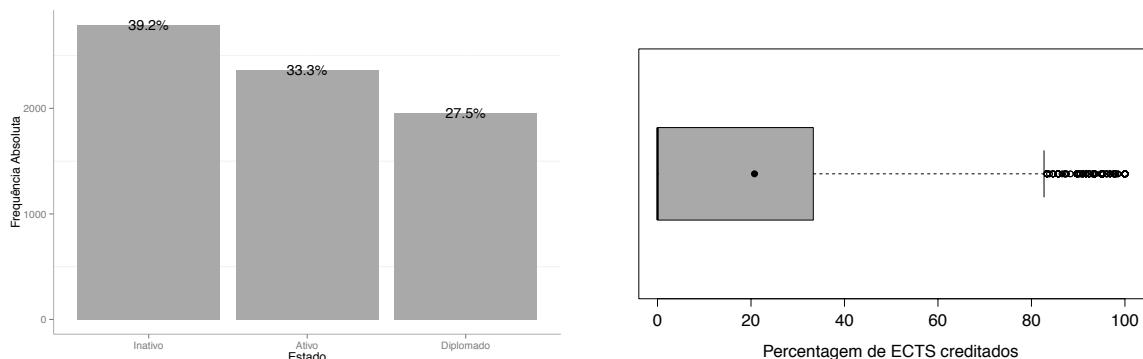
Cerca de 52,5% dos alunos ingressou na Universidade até ao ano letivo de 2009/2010 e nos últimos anos letivos registou-se um decréscimo do número de alunos com ingresso em cursos de Mestrado ou Doutoramento (Figura 2.2). Por outro lado, verificamos que a esmagadora maioria destes alunos ingressou na Universidade de Évora em cursos de Mestrado (84,7%) e em cursos pertencentes à Escola de Ciências Sociais (56,8%) (Figura 2.2).

Até à data de recolha dos dados, apenas 27,5% dos alunos terminou o seu curso, sendo que quase 40% dos alunos estava inativo (Figura 2.3). A esmagadora maioria destes alunos não pediu reingressos (96,3%) nem semestres adicionais (79,1%) e 50% dos alunos não obtiveram qualquer creditação de ECTS (Figura 2.3). Note-se que, devido a alguns valores omissos, apenas pudemos considerar a percentagem de ECTS creditados de 7096 alunos. Importa também referir que optámos por recorrer à percentagem de ECTS creditados e não ao número de ECTS, pois nem todos os cursos têm o mesmo número de ECTS para a parte curricular.

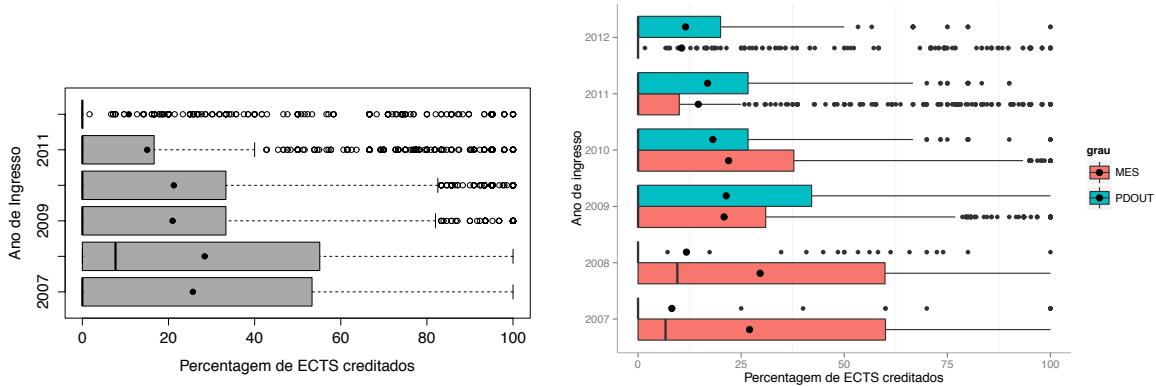
A distribuição da percentagem de créditos pelos diferentes anos de ingresso (Figura 2.4) não é homogénea. Assim, apenas destacamos que as medianas da percentagem de ECTS creditados são mais elevadas nos anos de ingresso de 2007 e 2008, 7% e 10%, respetivamente. Para os cursos de Doutoramento, concluímos que seja qual for o ano de ingresso considerado, metade dos alunos não obteve qualquer percentagem de ECTS creditados (Figura 2.4).



**Figura 2.2: Ano de ingresso (7116 alunos); Escola do curso de Mestrado ou Doutoramento (7116 alunos).**



**Figura 2.3: Estado do aluno (7116 alunos); Distribuição da percentagem de ECTS creditados aquando do ingresso na Universidade (7096 alunos).**



**Figura 2.4: Distribuição da percentagem de ECTS creditados por ano de ingresso (7096 alunos); Distribuição da percentagem de ECTS creditados por ano de ingresso e pelo grau do curso (7096 alunos).**

Para os estudos efetuados, analisámos subamostras de dimensões diferentes devido a algumas características das variáveis resposta em causa. O tempo até à entrega de uma dissertação ( $T_{diss}$ ), sobre o qual recorremos à análise de sobrevivência, foi definido como o tempo que decorre entre a data de aprovação do projeto e a data de entrega da dissertação. Neste caso, para alguns alunos, não existia informação sobre a data de aprovação do projeto, por isso apenas pudemos considerar 2164 alunos com registo desta data para os cursos de Mestrado e 158 alunos para os cursos de Doutoramento.

Para o estudo do perfil do aluno que pede um ou mais semestres adicionais num curso de Mestrado, apenas considerámos os 1819 alunos diplomados que estavam em condições de efetuar este pedido. No caso do estudo dos tempos de procedimentos que decorrem até as provas públicas, também não foi possível obter a informação da data de entrega da dissertação de Mestrado de todos os alunos. Deste modo, para aplicar os métodos de controlo estatístico, apenas considerámos 1095 alunos diplomados. Quanto ao tempo de espera pela emissão de um diploma, considerámos 692 alunos para o diploma de curso e 3732 alunos para o diploma de grau.

Tendo em conta algumas das particularidades das diferentes subamostras, optámos por aprofundar a caracterização das mesmas nas respetivas secções.

## II.2. Metodologia aplicada

Para modelar o tempo até à conclusão de uma dissertação de Mestrado recorreu-se à Análise de Sobrevivência. Esta pode ser definida por um conjunto de processos estatísticos para a análise de dados cuja variável resposta de interesse é o tempo até à ocorrência de um evento que

pode ou não ter ocorrido no final do período de estudo. Este tempo pode ser definido em anos, meses, semanas ou dias e o evento pode ser a morte de um paciente, a incidência de uma doença, a resposta a um tratamento, a remissão de uma doença, o tempo de vida de um equipamento eletrónico ou, neste caso particular, a conclusão de uma dissertação. (Kleinbaum, 2012)

Atendendo a que estamos na presença de dados censurados, ou seja, alunos ativos e inativos que estavam a realizar a sua dissertação no final do período de *follow-up* e para os quais desconhecemos o tempo até à ocorrência do evento, recorreu-se a um modelo de Cox. Com este modelo pretendeu-se encontrar fatores importantes na duração de uma dissertação de Mestrado. Neste estudo, considerou-se como variáveis explicativas a idade, sexo, nacionalidade, nível de escolaridade dos pais, tipologia e distrito da área de residência, ano de ingresso na Universidade, modo de ingresso, número de anos que o aluno esteve sem estudar até ingressar no curso, fase de candidatura, escola do curso, percentagem de ECTS creditados, existência do estatuto de trabalhador estudante, número de reingressos, tempo que o aluno levou a terminar a parte curricular do seu curso e média da parte curricular do curso. Para os tempos até à conclusão de uma dissertação de Doutoramento, considerou-se a mesma metodologia e as mesmas variáveis explicativas.

Para traçar o perfil mais provável do aluno que necessita pedir pelo menos um semestre adicional para terminar a sua dissertação de Mestrado, recorreu-se a um modelo de regressão logística. Segundo Hosmer e Lemeshow (2013), os métodos de regressão tornaram-se uma componente integral de qualquer análise de dados que procura descrever a relação entre uma variável resposta e uma ou mais variáveis explicativas. Contudo, quando estamos na presença de variáveis resposta dicotómicas, como é o caso da variável de interesse deste estudo (0 – não pediu semestres adicionais; 1 – pediu pelo menos um semestre adicional), os modelos de regressão logística são frequentemente utilizados. Assim, para identificar determinantes da decisão de pedir pelo menos um semestre adicional, ajustou-se um modelo de regressão logística, considerando como variáveis explicativas a idade, sexo, nacionalidade, nível de escolaridade dos pais, tipologia e distrito da área de residência, ano de ingresso na Universidade, fase de candidatura, número de anos que o aluno esteve sem estudar até ingressar no curso, escola do curso, percentagem de ECTS creditados e existência do estatuto de trabalhador estudante.

Segundo Montgomery (2009), os processos do controlo estatístico (SPC – *Statistical Process Control*) podem ser aplicados a uma diversidade de estudos. No caso particular desta dissertação, recorreu-se às cartas de controlo e à análise da capacidade dos processos relacionados com os procedimentos administrativos da Universidade. Neste sentido, tendo em conta a distribuição do

tempo que decorre entre a entrega da dissertação de Mestrado e a realização de provas públicas, implementaram-se cartas de controlo EWMA para amostras de dimensão 1. Depois, para compreender melhor o processo e para tentar detetar causas responsáveis pela variação do mesmo, estratificou-se o referido tempo pelos diferentes anos de conclusão, escolas, departamentos e, eventualmente, por alguns cursos de Mestrado.

Para os restantes tempos dos procedimentos administrativos verificou-se que os tempos analisados são muito assimétricos e com caudas pesadas. Tendo em conta que nestas condições não é recomendável a aplicação de cartas baseadas na distribuição normal, implementaram-se cartas de controlo X construídas com base em distribuições normais assimétricas, definidas em Figueiredo e Gomes (2013).

### **II.3. Pesquisa Bibliográfica**

Segundo Kleinbaum (2012) a análise de sobrevivência tem como objetivo estudar o tempo até à ocorrência de um determinado evento, sendo que este evento pode ser definido como qualquer experiência que ocorre a um indivíduo num determinado período de tempo. No caso particular desta dissertação, o evento de interesse é a entrega de uma dissertação de Mestrado ou Doutoramento. Recentemente, esta temática foi abordada por Teixeira *et al.* (2012), onde os autores modelaram o tempo até à conclusão de um curso superior de 1º Ciclo considerando uma amostra dos alunos que ingressaram na Universidade de Évora entre os anos letivos de 2005/2006 e 2009/2010. Neste estudo, os autores recorreram a estimativas de Kaplan-Meier e a um modelo de Cox para modelar o tempo até à conclusão do curso, utilizando como variáveis explicativas a idade, sexo, fase de candidatura, modo de ingresso, escola do curso de 1º Ciclo, opção de candidatura, nota de ingresso, estado do aluno, distrito e tipologia da área de residência, nacionalidade e nível de escolaridade dos pais. Com este estudo, os autores concluíram que os alunos do sexo masculino com mais de 18 anos e que ingressaram em cursos das escolas de Ciências e Tecnologia ou de Artes, têm um risco muito elevado de demorar mais tempo a concluir a sua Licenciatura.

Também Júnior *et al.* (2012) recorreram a esta metodologia para estudar o tempo de permanência dos estudantes num curso de Física numa Universidade do Brasil, considerando uma amostra de 1447 alunos que ingressaram na Universidade no período de 1995 a 2009. Estes autores utilizaram a análise de sobrevivência para compreender melhor o fluxo escolar dos alunos na sua Universidade e recorreram a variáveis relativas ao estado do aluno, média de ingresso e sexo do

aluno. Ao utilizar as estimativas de Kaplan-Meier, os autores concluíram que há uma maior probabilidade dos alunos concluir o curso entre o 3º e o 4º ano. Após esse período, a probabilidade dos alunos terminarem o curso reduz, tornando-se praticamente nula para os estudantes com mais de 8 anos de permanência na Universidade. Ao utilizar um modelo de Cox, os autores concluíram que os alunos que ingressaram na Universidade com médias elevadas apresentam um risco maior de terminarem o seu curso mais cedo.

Min *et al.* (2011) consideraram a mesma abordagem, contudo aplicaram a análise de sobrevivência ao tempo que decorre até à desistência de um curso de engenharia e não até à conclusão do curso. Apesar do evento de interesse ser diferente, a abordagem destes autores é muito semelhante. Os autores seguiram um grupo de 100179 alunos de várias Universidades de Engenharia nos Estados Unidos da América por um determinado período de tempo (entre 1987 e 2005) e consideraram como variáveis explicativas a idade, sexo e raça, variáveis relacionadas com o *background* familiar, o tipo de instituição, a dimensão da instituição, a média final do secundário e a classificação dos SAT<sup>1</sup>. Os autores compararam, através de testes apropriados, as curvas de sobrevivência de Kaplan-Meier de diferentes grupos de alunos em função do semestre. Com esta abordagem, concluíram que nos semestres iniciais, mulheres e caucasianos têm uma maior probabilidade de desistir dos seus cursos. Também verificaram que os alunos com classificações baixas nos SAT apresentam maiores probabilidades de desistir do curso durante o segundo semestre.

Na Europa, também se tem verificado alguma preocupação sobre a desistência de cursos Superiores. Gury (2011) modelou o tempo até um aluno desistir do curso superior, com o objetivo de proporcionar ações que possam melhorar a eficiência das Universidades. O autor considerou uma base de dados com o registo de 27000 alunos que ingressaram nas Universidades Francesas desde o ano de 1989 e concluiu que o facto dos alunos receberem o apoio financeiro dos pais diminui o risco de desistirem do curso mais cedo. Por outro lado, concluiu que o facto do aluno não ter ingressado na Universidade na primeira opção de candidatura e o facto do nível de escolaridade do pai ser inferior ao secundário, aumentam o risco dos alunos desistirem do curso mais cedo.

Relativamente a cursos de Doutoramento, Jones-White *et al.* (2013) recorreram à análise de sobrevivência, nomeadamente a estimativas de Kaplan-Meier e a um modelo de Cox estratificado para modelar o tempo até à conclusão de um curso de Doutoramento na Universidade

---

<sup>1</sup> SAT – Teste de aptidão utilizado na admissão às Universidades dos Estados Unidos da América. Estes testes foram introduzidos em 1926 com o nome *Scholastic Aptitude Test*. Atualmente, são designados apenas por SAT e compostos por duas categorias: SAT math e SAT verbal.

do Minnesota, nos EUA. Os autores consideraram, para além da idade de ingresso na Universidade, o sexo, a raça e a nacionalidade, variáveis como a existência de filhos e o estado civil. Para além disso, tiveram acesso a variáveis relacionadas com as características do programa de Doutoramento, com a performance dos alunos (número de apresentações ou publicações), com os recursos e suporte financeiro (bolsas de investigação ou empréstimos de graduação) e com fatores de satisfação (satisfação com o programa ou com o(s) orientador(es)). Com este estudo, os autores concluíram que os cursos com programas mais pequenos e com melhores relações entre aluno/universidade conduzem a menores tempos de conclusão dos cursos. Também concluíram que as bolsas facultadas aos alunos, que permitem reduzir os empréstimos de graduação, assim como a disponibilidade dos orientadores, têm um impacto positivo no tempo até à conclusão do curso. Outra conclusão relevante é que os alunos que têm uma maior experiência em publicações e comunicações apresentam um maior risco de terminar o seu curso mais cedo. Quanto às características individuais dos alunos, os autores apenas concluíram que os alunos que não têm filhos apresentam um maior risco de terminar o curso mais cedo do que aqueles que têm pelo menos um filho.

Em estudos recentes sobre o tempo de duração dos cursos de Doutoramento, nem todos os autores utilizam modelos de regressão de Cox ou estimativas de Kaplan-Meier. Em alguns dos estudos efetuados, os autores apenas consideram os dados relativos aos alunos que terminaram o seu Doutoramento, modelando a variável resposta através de modelos de regressão linear múltipla (Adebi, 1987) ou de modelos de risco de Weibull (Siegfried, 1999). No primeiro caso, os autores utilizaram uma amostra de 4255 alunos que ingressaram na Universidade UCLA<sup>2</sup> entre 1976 e 1985, concluindo que os alunos que usufruíam de bolsas ou outro tipo de financiamento apresentavam médias do tempo de conclusão do seu curso consideravelmente inferiores. Os autores também concluíram que os homens demoram, em média, menos 1,2 anos a terminar o curso do que as mulheres e que à medida que decresce o número de dependentes a cargo do aluno, diminui também o tempo até à conclusão do Doutoramento. Já no estudo de Siegfried (1999), os autores recorreram a uma amostra de 618 alunos de economia dos EUA que completaram o seu Doutoramento entre 1996 e 1997. Concluíram que aqueles alunos que usufruem de bolsas ou outro tipo de financiamento terminam o seu curso mais cedo do que os alunos que não têm qualquer tipo de financiamento ou que, para além de serem alunos de Doutoramento, são também professores assistentes.

---

<sup>2</sup> UCLA – Universidade da Califórnia, Los Angeles – EUA.

Outra abordagem foi efetuada pela Agência de Estatística da Educação Superior do Reino Unido (HEFCE, 2005), na qual foram aplicados modelos de regressão logística para estudar os fatores determinantes para a não conclusão de um curso superior no tempo previsto. Este estudo foi realizado a partir de uma amostra 19191 alunos que ingressaram em Universidades do Reino Unido entre 1995 e 2002. Também estes autores concluíram que aqueles alunos que usufruem de bolsas ou outro tipo de financiamento têm mais possibilidades de completar o seu curso no tempo previsto.

Apesar de encontrarmos alguns estudos sobre esta temática, sabemos que a realidade das Universidades Portuguesas, seja nos cursos de Mestrado ou de Doutoramento, não é a mesma das Universidades Americanas ou Brasileiras. Para além disso, na Europa existem poucos estudos sobre esta temática e muitos dos estudos estão mais focados na problemática das desistências dos cursos superiores. Em nenhum dos estudos mencionados os autores abordam esta temática segundo a nossa perspetiva, i.e., isolando o efeito que os procedimentos administrativos podem ter no tempo que decorre até à conclusão do curso. Esta situação, só vem dificultar ainda mais uma possível comparação entre as abordagens efetuadas. Porém, na generalidade dos estudos relativos aos cursos de Doutoramento, podemos concluir que o financiamento proporcionado aos alunos é um fator determinante tanto no tempo até à conclusão do curso como nas possibilidades do aluno terminar ou não o seu curso. Note-se que no caso da nossa amostra não tivemos acesso a esta informação, nem às informações relacionadas com o grau de satisfação dos alunos relativamente ao apoio dos diferentes departamentos e dos orientadores associados à Universidade de Évora.

Também destacamos que a média de ingresso ou a opção de candidatura são fatores comuns a alguns dos estudos mencionados, seja os que consideram o tempo até à conclusão ou os que abordam o tema da desistência de um curso superior.

Os modelos de regressão logística também têm sido utilizados nos estudos relativos às problemáticas do ensino superior. Contudo, não encontrámos qualquer publicação da utilização destes modelos na identificação de fatores que levam um aluno a pedir semestres adicionais. O pedido de um semestre adicional não é comum a todas as Universidades da Europa, por isso seria de esperar encontrar alguma dificuldade na pesquisa deste tema tão específico. Apesar das dificuldades encontradas, verificámos que no contexto do ensino superior existem alguns autores que recorreram a esta metodologia, em particular para estudar determinantes para a conclusão de um curso superior.

Para além do estudo realizado pela HEFCE (2005), destacamos o trabalho de Hosoi *et al.* (2011). Estes autores realçam a problemática das diferenças entre os homens e mulheres no ingresso e conclusão de um curso superior de Engenharia. Através de uma amostra de 470 alunos que ingressaram na Universidade CSU<sup>3</sup> em 2005, os autores concluíram que os alunos que são casados, que ingressam na Universidade na primeira fase e com médias mais elevadas têm mais possibilidades de concluir o curso de engenharia. Concluíram também que as mulheres de nacionalidade americana têm mais possibilidades de terminar o curso do que os homens americanos.

Apesar da grande aplicabilidade dos processos de controlo estatístico, não encontrámos estudos que relacionam esta temática com os processos administrativos das Universidades. Para além disso, desconhecemos qualquer aplicação das cartas X adaptadas a distribuições *Skew-Normal* associadas à temática do ensino. Contudo, destacamos o estudo efetuado recentemente por Green Jr. *et al.* (2012), onde os autores propõem a implementação de cartas de controlo estatístico no processo de monitorização do serviço educativo das escolas de Gestão. Para além disso, consideram que o seu estudo oferece uma abordagem diferente que permite uma melhoria contínua dos programas das unidades curriculares dos cursos de Gestão. Neste sentido, desenvolveram cartas de controlo para monitorizar a média dos exames dos alunos e a proporção de alunos que assiste às aulas. Esta abordagem tem mais uma particularidade, na medida em que este processo é implementado no início de cada ano letivo com a participação dos alunos, seja na construção das cartas de controlo como na sua monitorização até ao final do período escolar. Ao contrário do estudo por nós elaborado, os autores consideraram amostras com dimensões superiores a um, visto que cada amostra representa uma turma com mais alunos. Os autores concluem que este tipo de monitorização, para além de ser benéfica para a avaliação da *performance* dos alunos ao longo do tempo é também uma ferramenta extremamente útil para que os alunos possam atingir os objetivos propostos nas unidades curriculares dos cursos de Gestão.

---

<sup>3</sup> CSU – Colorado State University – EUA.



## **Capítulo III – Modelação do tempo até à conclusão de uma dissertação**

### **III.1. Introdução**

Segundo Quinn (2013), são demasiados os estudantes da União Europeia que abandonam os estudos antes de terminarem o seu curso superior. Para além disso, o facto de alguns alunos prolongarem a sua estadia na Universidade pode ter impacto não só nos orçamentos das famílias e das Universidades, mas também no crescimento económico do país.

Neste capítulo, pretendemos encontrar os fatores que têm uma influência determinante neste adiamento, pois a compreensão do comportamento dos alunos pode ser importante na conceção de políticas de intervenção e de apoio aos alunos desta Universidade. Para tal, tivemos em conta apenas o desempenho dos alunos e removemos o efeito que os procedimentos administrativos da Universidade podem ter na conclusão de um curso.

Para a realização deste estudo, recorremos à Análise de Sobrevida, mais precisamente, a um modelo semi-paramétrico de riscos proporcionais de Cox. Com esta abordagem pretendemos estudar o tempo até à conclusão de uma dissertação de Mestrado e, posteriormente, de Doutoramento, estimando a razão de riscos de um grupo de indivíduos com determinadas características em relação a outro.

### **III.2. Metodologia aplicada**

Na análise de sobrevida, normalmente, referimo-nos à variável resposta como o tempo de sobrevida, ou seja, o tempo ao fim do qual um indivíduo sobreviveu a um evento. O evento pode ser a morte de um paciente, a incidência ou a recuperação de uma doença. Porém, na maioria destes estudos, quando termina o período de *follow-up* nem todos os indivíduos experienciaram o evento e, apesar de termos alguma informação sobre estes indivíduos, não sabemos ao certo qual o seu tempo de sobrevida. Estas observações são consideradas censuradas e, apesar de fornecerem informações incompletas, não devem ser descartadas, pois a sua exclusão da análise pode introduzir algum enviesamento nas funções de sobrevida. (Kleinbaum, 2012)

As censuras podem ocorrer quando um indivíduo não experienciou o evento até ao final do período de estudo ou quando o indivíduo desiste do estudo. Neste caso particular, o evento é a entrega da dissertação e os tempos censurados são relativos àqueles alunos que inativaram a sua

inscrição (alunos inativos) ou aos alunos que não entregaram a sua dissertação até ao final do período de *follow-up* (alunos ativos).

Para o caso dos alunos que experienciaram o evento, a variável resposta foi definida como o tempo que decorre entre a data de aprovação do projeto pela comissão de curso e a data de entrega da dissertação. Mas, para aqueles alunos que precisaram de reformular as suas teses, foi necessário acrescentar o tempo de reformulação da tese. Este tempo foi definido como o tempo que decorre entre a data de receção do despacho liminar e a data de entrega da tese após reformulação (Figura 1.1, pág. 4). Para outros alunos, também tivemos de considerar uma ou mais inativações da sua inscrição. Para tal, ao tempo de dissertação tivemos de retirar o(s) tempo(s) até ao reingresso na Universidade. Definimos este tempo como o tempo que decorre entre a data de inativação e a data de reingresso na Universidade.

No caso das observações censuradas, o cálculo do tempo de dissertação foi efetuado de duas formas diferentes. Para o caso dos alunos inativos, considerámos o tempo que decorre entre a data de aprovação do projeto e a data de inativação e, quando necessário, também acrescentámos o tempo de reformulação ou retirámos o tempo de outras inativações. Já no caso dos alunos ativos, calculámos o tempo que decorre entre a data de aprovação do projeto e a data de recolha dos dados, sendo que também neste caso, para alguns alunos foi preciso acrescentar o tempo de reformulação da tese ou o(s) tempo(s) de inativação.

Na análise de sobrevivência existem diversas formas de especificar a distribuição do tempo de vida dos indivíduos. A função de sobrevivência  $S(t)$  é a probabilidade que um indivíduo tem de sobreviver ao evento até um determinado tempo  $t$ , isto é,  $S(t)$  dá-nos a probabilidade da variável aleatória  $T$  exceder um tempo específico  $t$

$$S(t) = P(T > t). \quad (3.1)$$

Por outro lado, define-se a distribuição do tempo de vida  $F(t)$  como a probabilidade do evento ocorrer até ao tempo  $t$

$$F(t) = P(T \leq t), \quad (3.2)$$

tal que  $F(t) + S(t) = 1$ , em qualquer tempo  $t$ .

Definimos a função de risco  $h(t)$  como a probabilidade de um determinado indivíduo experienciar o evento no intervalo de tempo instantâneo ( $t \leq T < t + \Delta t$ ), no universo daqueles que sobreviveram até ao tempo  $t$ , por unidade de tempo  $\Delta t$ . Assim, a função de risco é definida

como uma taxa de risco instantânea da ocorrência do evento no instante  $t$ , condicional à sobrevivência até esse instante (Kleinbaum, 2012),

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}, \text{ tal que: } 0 \leq h(t) < \infty. \quad (3.3)$$

A função de risco pode apresentar várias formas:

1. Pode ser constante ao longo do tempo, tal que  $h(t) = \lambda$ . Neste caso, consideramos o modelo de sobrevivência exponencial.
2. Pode aumentar ou diminuir ao longo do tempo, sendo que um modelo de *weibull* crescente ou decrescente pode ser adequado a este tipo de funções.
3. Pode ser crescente nos primeiros períodos de tempo e decrescente ao fim de um determinado tempo. Neste caso, um modelo *log-normal* pode ser o mais adequado.

Note-se que a função de risco foca-se na ocorrência do evento, ao contrário da função de sobrevivência que se foca na não ocorrência do evento. Efetivamente, as informações dadas por estas funções baseiam-se em conceitos opostos e é possível definir uma em função da outra. Se soubermos a forma de  $S(t)$ , é possível obter a função de risco correspondente e vice-versa. Por exemplo, se  $h(t) = \lambda$  então a função de sobrevivência correspondente é dada por  $S(t) = e^{-\lambda t}$ . De uma forma geral:

$$S(t) = \exp \left[ - \int_0^t h(u) du \right] \quad (3.4)$$

$$h(t) = - \left[ \frac{dS(t) / dt}{S(t)} \right] \quad (3.5)$$

Por vezes, não é possível assumir uma distribuição conhecida para a função de risco, por isso a maioria dos estudos são efetuados tendo em conta outras técnicas de análise de sobrevivência, nomeadamente as técnicas não paramétricas e semi-paramétricas.

As técnicas não paramétricas não apresentam restrições sobre a distribuição dos tempos de vida das populações, contudo são um pouco limitadas, pois não permitem testar o efeito de diferentes covariáveis ao mesmo tempo. A técnica principal não paramétrica da análise de sobrevivência tem por base o estimador de Kaplan-Meier, dado por

$$\hat{S}_{KM}(t) = \hat{S}(t_{(f)}) = \hat{S}(t_{(f-1)}) \times \hat{P}(T > t_{(f)} | T \geq t_{(f)}) = \prod_{i=1}^f \hat{P}(T > t_{(i)} | T \geq t_{(i)}) \quad (3.6)$$

e que traduz a probabilidade do indivíduo sobreviver ao tempo em que ocorre o evento  $f$ ,  $t_{(f)}$ . Este estimador pode ser definido como a probabilidade de sobreviver ao tempo do evento anterior,  $\hat{S}(t_{(f-1)})$ , multiplicada pela probabilidade de sobreviver ao tempo  $t_{(f)}$ , sabendo que sobreviveu pelo menos ao tempo  $t_{(f)}$  (Kleinbaum, 2012).

O intervalo de confiança a 95% para as estimativas de KM pode ser obtido através da expressão:

$$\hat{S}_{KM}(t) \pm 1,96 \sqrt{V \hat{\text{ar}}[\hat{S}_{KM}(t)]}, \quad (3.7)$$

$$\text{onde } V \hat{\text{ar}}[\hat{S}_{KM}(t)] = \left( \hat{S}_{KM}(t) \right)^2 \sum_{f: t_{(f)} \leq t} \left[ \frac{m_f}{n_f(n_f - m_f)} \right],$$

$t_{(f)}$  = tempo até à ocorrência do evento  $f$ ,

$m_f$  = número de eventos que ocorreram no  $t_{(f)}$  e

$n_f$  = número de indivíduos em risco no  $t_{(f)}$ .

Através das estimativas de Kaplan-Meier, é possível construir as curvas que traduzem a probabilidade de sobrevivência em cada um dos tempos  $t_{(f)}$ . As curvas de Kaplan-Meier são construídas de acordo com a ordem pela qual ocorrem os diferentes eventos e apresentam “degraus” à medida que avançamos de evento para evento. Se, por exemplo, quisermos comparar as probabilidades de sobrevivência entre os homens e as mulheres em cada um dos tempos  $t_{(f)}$ , as curvas de KM são uma ferramenta importante.

Para comparar as curvas de KM de dois ou mais grupos de indivíduos com características diferentes, recorre-se a testes adequados para avaliar se as curvas são estatisticamente equivalentes.

Quando as curvas de KM não se intersectam, um dos testes mais utilizados é o teste log-rank, quando se intersectam o teste de Peto é o mais adequado. A fórmula da estatística do teste do teste de log-rank é obtida pela expressão

$$\chi^2 \approx \sum_i^{n^{\text{grupos}}} \frac{(O_i - E_i)^2}{E_i} \quad (3.8)$$

e representa uma aproximação do teste clássico do qui-quadrado, onde  $O_i - E_i$  é a diferença entre os valores observados e esperados para cada grupo de indivíduos  $i$  (Kleinbaum, 2012).

O teste de Peto é mais robusto quando as curvas de KM se intersectam. A estatística de teste é uma variação da expressão (3.8), à qual se aplica um peso  $w(t_{(f)})$  a cada tempo que ocorre o evento  $f$ , calculado pela sobrevivência estimada,  $\tilde{s}(t_{(f)})$ , para o conjunto de todos os grupos de indivíduos (Kleinbaum, 2012).

Uma das técnicas semi-paramétricas mais utilizada é o ajustamento de um modelo semi-paramétrico de riscos proporcionais de Cox. Este modelo tem a particularidade de permitir estudar o tempo até à ocorrência do evento incluindo a informação de mais do que uma variável explicativa, sem ter de especificar qual é a distribuição do tempo de vida da população em estudo. O modelo de regressão de Cox pode ser definido pela expressão

$$h(t, \mathbf{X}) = h_0(t) \exp(x_1 \beta_1 + x_2 \beta_2 + \dots + x_p \beta_p) = h_0(t) \exp(\mathbf{X} \mathbf{B}), \quad (3.9)$$

tal que  $h_0(t)$  é a função de risco basal,  $\mathbf{X}$  é o vetor das  $p$  variáveis explicativas e  $\mathbf{B}$  o vetor dos coeficientes  $\beta_1, \beta_2, \dots, \beta_p$ .

A medição do efeito das variáveis explicativas é efetuada estimando a razão de riscos (HR) de um determinado grupo de indivíduos em relação a outro, tal que:

$$\hat{HR} = \frac{\hat{h}(t | \mathbf{X}^*)}{\hat{h}(t | \mathbf{X})} = \frac{e^{\sum_{i=1}^p \hat{\beta}_i X_i^*}}{e^{\sum_{i=1}^p \hat{\beta}_i X_i}} = \exp \left[ \sum_{i=1}^p \beta_i (X_i^* - X_i) \right], \quad (3.10)$$

onde  $\mathbf{X} = (X_1, X_2, \dots, X_p)$  e  $\mathbf{X}^* = (X_1^*, X_2^*, \dots, X_p^*)$  denotam os vetores das covariáveis de dois grupos de indivíduos distintos.  $X_i^* = 1$  define os indivíduos expostos a determinadas características das variáveis  $i$  e  $X_i = 0$  os indivíduos não expostos a essas características (Kleinbaum, 2012).

Por exemplo, se quisermos estimar a razão de riscos dos homens em relação à mulheres, então  $\mathbf{X} = (X_1, X_2, \dots, X_p) = X_1$  e  $\mathbf{X}^* = (X_1^*, X_2^*, \dots, X_p^*) = X_1^*$ , onde  $X_1^*$  toma o valor 1 para o grupo dos homens e  $X_1$  toma o valor 0 para o grupo das mulheres, tal que:

$$\hat{HR} = \exp\left[\hat{\beta}_1(X_1^* - X_1)\right] = \exp\left[\hat{\beta}_1(1 - 0)\right] = e^{\hat{\beta}_1} \text{ e } IC_{95\%}(\hat{HR}) = \exp\left[\hat{\beta}_1 \pm 1,96\sqrt{Var(\hat{\beta}_1)}\right].$$

De uma forma geral, o intervalo de confiança associado à razão de riscos entre dois grupos (0,1) de uma variável  $X_1$  com interações  $X_1 \times W_1, \dots, X_1 \times W_k$  é calculado pela seguinte expressão:

$$IC_{95\%}(\hat{HR}) = \exp\left[\hat{l} \pm 1,96\sqrt{Var(\hat{l})}\right], \quad (3.11)$$

onde  $\hat{l} = \beta_1 + \delta_1 W_1 + \delta_2 W_2 + \dots + \delta_k W_k$ ,

$\beta_1$  é o coeficiente de  $X_1$ ,

$\delta_j$  é o coeficiente de  $X_1 \times W_j, j = 1, \dots, k$  e

$$Var(\hat{l}) = Var(\hat{\beta}_1) + \sum_j W_j^2 Var(\hat{\delta}_j) + 2 \sum_j W_j Cov(\hat{\beta}_1, \hat{\delta}_j) + 2 \sum_j \sum_k W_j W_k Cov(\hat{\delta}_j, \hat{\delta}_k).$$

Segundo Kleinbaum (2012), tipicamente, a formulação da função de verosimilhança é baseada na distribuição da variável resposta. Porém, no caso do modelo de Cox a função de verosimilhança é baseada na ordem dos eventos observados e não na distribuição conjunta dos eventos. Por isso, neste caso estamos perante uma função de verosimilhança parcial  $L$ , dada pelo produto das verosimilhanças obtidas para cada tempo em que ocorreu o evento  $f$

$$L = L_1 \times L_2 \times \dots \times L_k = \prod_{f=1}^k \left( \frac{\exp\{\mathbf{X}_f \mathbf{B}\}}{\sum_{j \in R(t_f)} \exp\{\mathbf{X}_f \mathbf{B}\}} \right)^{\delta_f}, \quad (3.12)$$

onde  $R(t_f)$  é o conjunto dos índices das observações sob risco no tempo  $t_{(f)}$  e  $\delta_f$  é a variável indicadora da ocorrência do evento  $f$ . As estimativas de máxima verosimilhança parcial do modelo de Cox,  $\hat{\beta}_i$ , são obtidas maximizando a função  $L$ .

A expressão do modelo de Cox (3.9) assume que as  $p$  variáveis explicativas têm um efeito multiplicativo na função de risco e, consequentemente, a razão entre o risco de ocorrência do

evento para dois indivíduos é constante no tempo. Deste modo, para que este modelo possa ser utilizado é necessário verificar o pressuposto de riscos proporcionais (RP). O pressuposto de RP pode ser verificado com uma abordagem gráfica (resíduos de Schoenfeld) ou através de testes adequados como o teste de Harrel (Kleinbaum, 2012).

Se estivermos perante  $p$  variáveis explicativas, ou  $p$  preditores, então teremos  $p$  resíduos de Schoenfeld definidos para cada indivíduo que experienciou o evento. Uma análise gráfica aos resíduos de Schoenfeld associados a cada covariável permite verificar se o comportamento dos resíduos ao longo do tempo é ou não o desejado. Por outro lado, o teste de Harrel permite testar se os resíduos de Schoenfeld associados a cada covariável estão correlacionados com o tempo (Kleinbaum, 2012).

Para avaliar a bondade do ajustamento dos modelos de Cox é possível utilizar uma medida análoga ao  $R^2$  de Nagelkerke

$$R_p^2 = 1 - \left\{ \exp \left[ \frac{2}{n} (L_0 - L_p) \right] \right\}, \quad (3.13)$$

onde  $n$  é a dimensão da amostra,  $L_p$  é a verosimilhança parcial para o modelo ajustado com  $p$  covariáveis e  $L_0$  a verosimilhança parcial para o modelo sem covariáveis.

Porém, Hosmer e Lemeshow (2008) referem que apesar de se poder recorrer a esta medida, esta pode ser baixa quando estamos perante uma amostra com muitos dados censurados e, consequentemente, traduzir uma medida pouco fiável para a qualidade do ajustamento do modelo.

Para ajustar um modelo multivariado de Cox, Hosmer e Lemeshow (2008) sugerem que se efetue os seguintes passos:

1. Para o modelo inicial, selecionar todas as variáveis que se revelam significativas na fase univariada (valores  $p$  do teste de Wald<sup>4</sup> inferiores a 0,20);
2. A partir desse modelo eliminar sucessivamente, e por ordem decrescente dos valores  $p$  de Wald, todas as variáveis não significativas ( $valor\ p > 0,05$ );
3. Verificar se as variáveis que não foram incluídas no modelo inicial se mostram significativas na presença das restantes, caso em que devem ser adicionadas ao modelo;

---

<sup>4</sup> O teste de Wald testa a hipótese  $H_0: \hat{\beta}_i = 0$ , para a covariável  $i$ , com base no desvio padrão estimado.

4. Ajustar as categorias das variáveis e verificar, para as variáveis contínuas, o pressuposto da linearidade com o logaritmo da função de risco, através do método dos quartis, resíduos Martingala e pelo método dos polinómios fracionários;
5. Testar, via teste de razão de verosimilhanças (TRV<sup>5</sup>), as interações entre as variáveis que fazem sentido no contexto do estudo (valor  $p \leq 0,05$ );
6. Através de uma análise de resíduos Deviance, Score e DfBetas, realizar uma análise a valores atípicos e observações influentes;
7. Validar o pressuposto dos riscos proporcionais, através do teste de Harrel e via análise de resíduos de Schoenfeld.
8. Avaliar o índice de prognóstico do modelo.

O índice de prognóstico do modelo é o preditor linear do modelo de Cox calculado para cada indivíduo, utilizando as covariáveis observadas e as estimativas dos coeficientes de regressão do modelo ajustado. Uma forma de avaliar este índice é através da comparação das funções de sobrevivência empírica e as funções de sobrevivência de KM para grupos de alto, médio e baixo índice de prognóstico.

### III.3. Caracterização da amostra

#### III.3.1. Cursos de Mestrado

Para os cursos de Mestrado, para construir um modelo de análise de sobrevivência, foram consideradas as variáveis da Tabela 3.1.

**Tabela 3.1: Variáveis consideradas para a análise de sobrevivência do tempo até à entrega de uma dissertação de Mestrado e respetivas categorias.**

Variável	Categoria
Sexo	Feminino
	Masculino
Nacionalidade	Portuguesa
	Outra
Idade	(anos)
Nível de Escolaridade dos pais	Ambos até ao ensino secundário
	Pelo menos um com ensino superior
NUTSII	Alentejo
	Centro
	Algarve

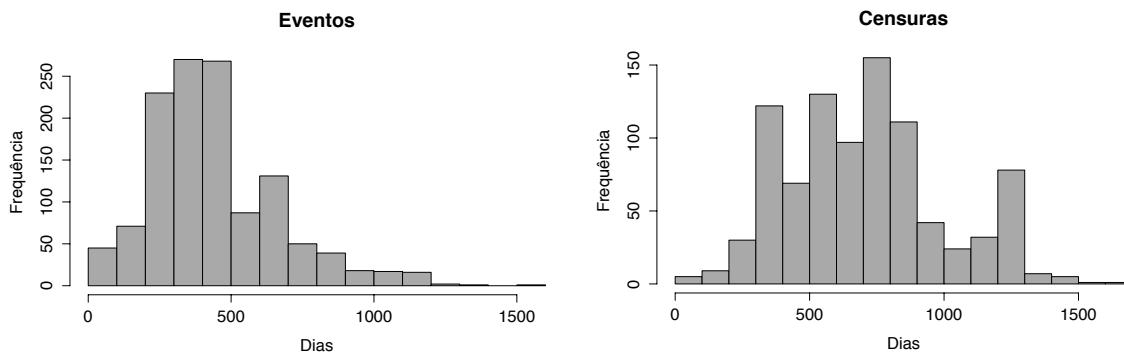
---

<sup>5</sup> O TRV testa a hipótese  $H_0: \hat{\beta}_i = 0$ , com base na razão entre as verosimilhanças do modelo com a covariável  $i$  e sem a covariável  $i$ .

Tipologia	Norte Açores Madeira APU AMU APR
Ano ingresso (AI)	2007 2008 2009 2010 2011
Fase de candidatura (FC)	1ª Fase 2ª Fase 3ª ou 4ª Fase
Escola do curso	Ciências Sociais (CS) Artes (A) Ciências e Tecnologia (CT)
Estatuto de trabalhador estudante (TE)	Não Sim
Número de reingressos	Zero Um ou mais
Percentagem de ECTS creditados	(%)
Duração da parte curricular (DC)	Terminou no tempo previsto (até 3 semestres) Não terminou no tempo previsto (mais de 3 semestres)

Foram considerados 2164 alunos com ingresso na Universidade de Évora entre os anos letivos de 2007/2008 e 2011/2012, tendo sido seguidos até Julho de 2013. Note-se que para este estudo não pudemos considerar os alunos com ingresso no ano letivo de 2012/2013, pois, à data da recolha dos dados, estes alunos ainda não estavam em condições de entregar a sua dissertação. Para além disso, não tivemos acesso à informação de todas as datas de aprovação do projeto de dissertação, o que resultou numa redução considerável da amostra inicial.

Concluímos que 57,6% dos alunos entregou a sua dissertação (eventos) e que as restantes observações (censuras) são relativas aos alunos para os quais o projeto de dissertação foi aceite pela comissão de curso, mas que, por alguma razão, não entregaram a sua dissertação final (alunos ativos e inativos) (Figura 3.1).



**Figura 3.1: Distribuição dos tempos até à entrega da dissertação de Mestrado (eventos); Distribuição dos tempos até à data de inativação ou de recolha dos dados (censuras).**

Dos alunos desta amostra, 58,6% são do sexo feminino e 76,6% são de nacionalidade Portuguesa. A idade de ingresso destes alunos está entre os 20 e 72 anos, sendo que 50% dos alunos ingressou na Universidade até aos 24 anos (Figura 3.2). A média é igual a 31,5 anos e o desvio padrão igual a 8,8 anos.

A maioria dos alunos ingressou na Universidade através de concurso local (99,8%), até ao ano letivo de 2008/2009 (53,1%) e apenas 4,4% dos alunos desta amostra ingressou no ano letivo de 2011/2012 (Figura 3.2). A escola das Ciências Sociais é a mais representada nos cursos de Mestrado, sendo que 59,3% dos alunos desta amostra ingressaram em cursos desta escola (Figura 3.3). Note-se que não há registo de cursos de Enfermagem, o que seria de esperar visto que apenas estamos a considerar os cursos cuja dissertação faz parte do programa. Também concluímos que a esmagadora maioria destes alunos não obteve o estatuto de trabalhador estudante (86,4%) nem fez qualquer reingresso na Universidade (96,4%).

Para algumas variáveis não obtivemos o registo de todos os alunos. Para as variáveis distrito e tipologia da área de residência apenas pudemos considerar a informação de 1657 alunos. Concluímos que a maioria destes alunos reside no Alentejo (52%) (Figura 3.3) e 74,4% dos alunos provêm de áreas predominantemente urbanas (APU), 12,5% de áreas medianamente urbanas (AMU) e 13,2% de áreas predominantemente rurais (APR).

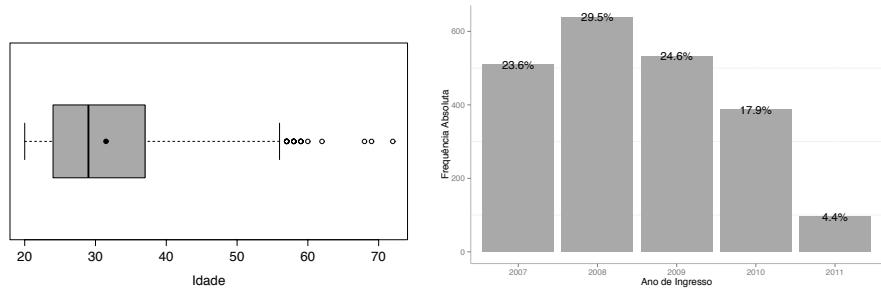


Figura 3.2: Distribuição das idades dos alunos de Mestrado (2164 alunos); Ano de ingresso (2164 alunos).

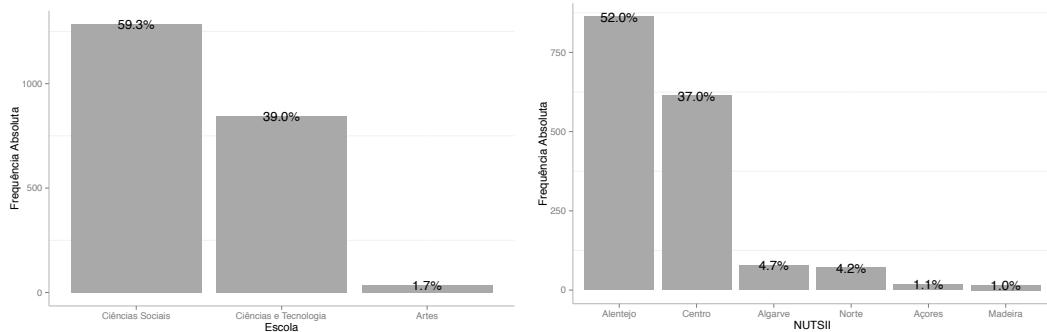


Figura 3.3: Escola do curso de Mestrado (2164 alunos); Distrito da área de residência (NUTSII) (1657 alunos).

No caso do nível de escolaridade dos pais, apenas tivemos acesso ao registo de 1667 alunos. Concluímos que para a maioria destes alunos, tanto as mães como os pais têm um nível de escolaridade até ao 2º Ciclo, 52,5% no caso das mães e 52,8% no caso dos pais (Figura 3.4). Também considerámos o nível de escolaridade dos pais agregando as variáveis anteriores numa variável com apenas duas categorias: 1 – ambos têm até ao ensino secundário; 2 – pelo menos um tem um nível de escolaridade superior. Concluímos que, para a maioria dos alunos, ambos os pais têm até ao ensino secundário (77,9%).

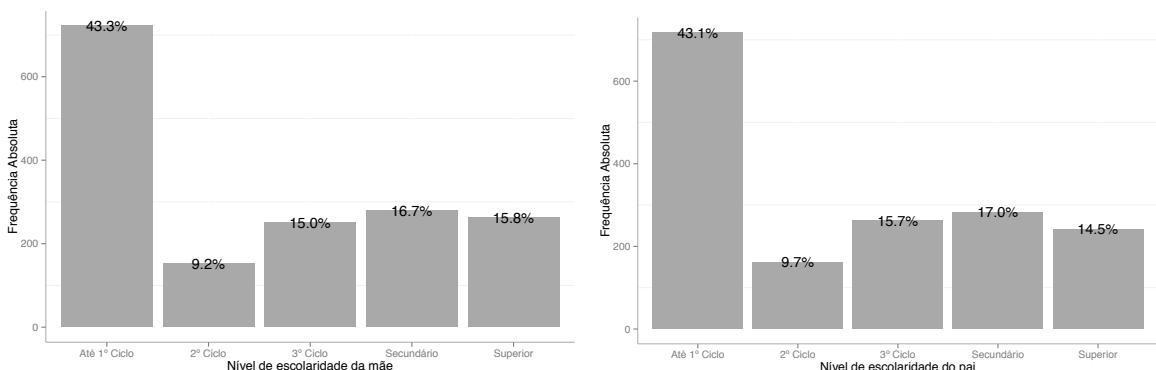
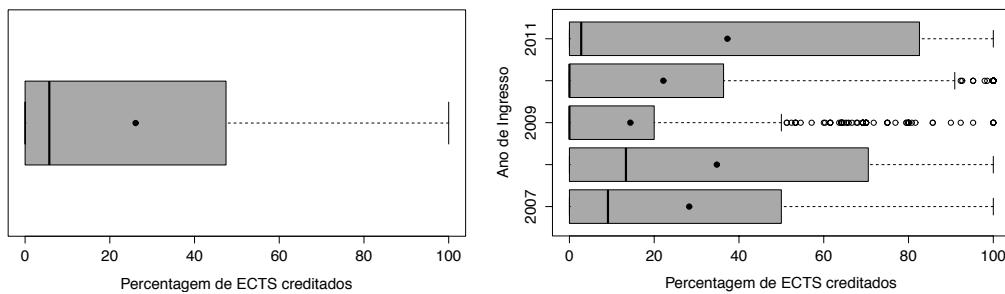


Figura 3.4: Nível de escolaridade da mãe (1667 alunos); Nível de escolaridade do pai (1667 alunos).

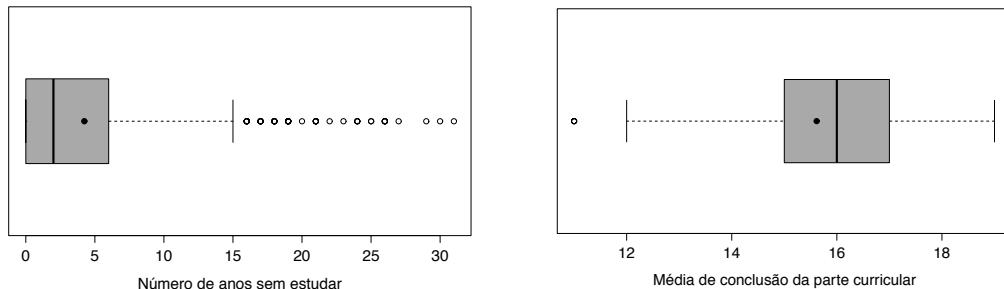
Também para as variáveis fase de candidatura, percentagem de ECTS creditados e duração da parte curricular apenas pudemos recorrer à informação de 1552, 2149 e 2010 alunos, respetivamente. Concluímos que a maioria dos 1552 alunos ingressou na primeira fase (84,5%), sendo que 9,5% ingressou na segunda fase e 5,9% na terceira ou quarta fase. A média da percentagem de ECTS creditados é igual a 26,1% e o desvio padrão igual a 35,2%, sendo que 50% dos alunos obteve até 6% de ECTS creditados (Figura 3.5). Salientamos que as médias mais elevadas são relativas aos anos de 2008 e 2011, 34,8% ( $sd=38,6$ ) e 37,3% ( $sd=42,3$ ), respetivamente.

No caso da duração da parte curricular, definimos esta variável como o tempo que decorre entre a data de ingresso dos alunos na Universidade e a data da última avaliação. Contudo, existiam alguns alunos que, apesar de estarem em condições de entregar as suas dissertações, ainda não tinham terminado a parte curricular aquando da data de recolha dos dados. Por isso, optou-se por categorizar esta variável usando como ponto de corte 3 semestres. Concluímos que dos 2010 alunos, 41% não terminou a parte curricular no tempo previsto, ou seja, estes alunos terminaram a parte curricular em mais de 3 semestres.



**Figura 3.5: Distribuição da percentagem de ECTS creditados (2149 alunos); Distribuição da percentagem de ECTS creditados por ano de ingresso (2149 alunos).**

Para as variáveis número de anos sem estudar e média da parte curricular apenas pudemos considerar 668 e 1133 alunos, respetivamente. Devido à redução considerável da amostra inicial, optámos por não recorrer a estas variáveis para o ajustamento do modelo de regressão de Cox. Porém, podemos concluir que 50% dos 668 alunos esteve até 2 anos sem estudar e que a média deste tempo é igual a 4,2 anos e o desvio padrão é igual a 5,9 anos (Figura 3.6). Por fim, concluímos que 50% dos 1133 alunos com registo da média de conclusão da parte curricular obteve até 16 valores, sendo que a média é igual a 15,6 valores e o desvio padrão é igual a 1,3 valores (Figura 3.6).



**Figura 3.6: Distribuição do número de anos sem estudar (668 alunos); Distribuição da média da parte curricular (1133 alunos).**

### III.3.2. Cursos de Doutoramento

Para os cursos de Doutoramento, para construir um modelo de análise de sobrevivência, foram consideradas as variáveis da Tabela 3.2.

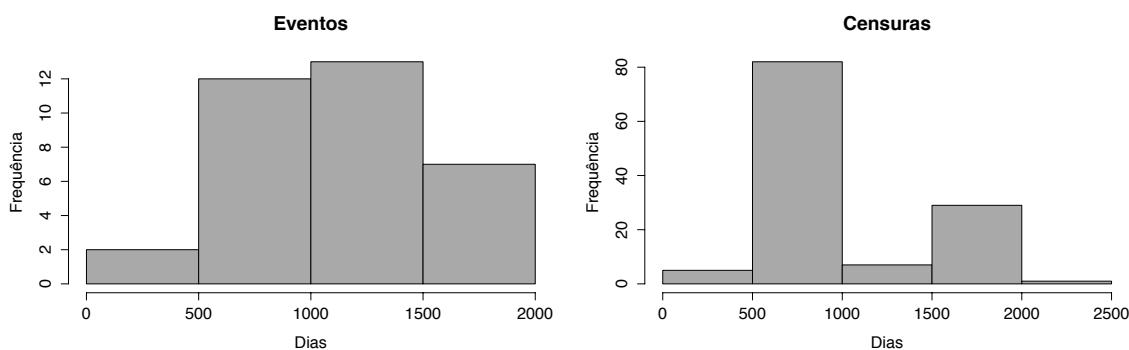
**Tabela 3.2: Variáveis consideradas para a análise de sobrevivência do tempo até à entrega de uma dissertação de Doutoramento e respetivas categorias.**

Variável	Categoría
Sexo	Feminino Masculino
Nacionalidade	Portuguesa Outra
Idade	(anos)
Nível de Escolaridade dos pais	Ambos até ao ensino secundário Pelo menos um com ensino superior
NUTSII	Alentejo Restantes
Tipologia	APU AMU APR
Ano ingresso	2007 2008 2009
Escola do curso	Ciências Sociais (CS) Artes (A) Ciências e Tecnologia (CT)
Percentagem de ECTS creditados	(%)
Duração da parte curricular <sup>1</sup>	Terminou no tempo previsto ou curso sem parte curricular Não terminou no tempo previsto

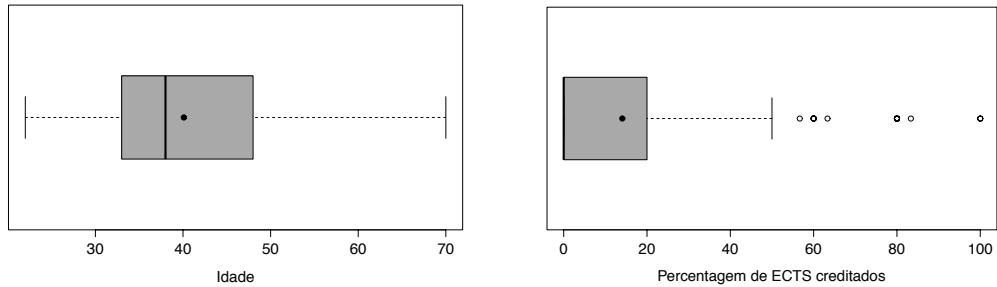
<sup>1</sup> Para os Doutoramentos o tempo previsto depende do curso, por isso não é possível definir um ponto de corte temporal.

A dimensão da amostra dos cursos de Doutoramento é igual 158 alunos que ingressaram na Universidade entre os anos letivos de 2007/2008 e 2009/2010. Neste caso, não considerámos os alunos com ingresso em 2010 e 2011, pois, à data da recolha dos dados, estes alunos ainda não estavam em condições de entregar a sua dissertação. Esta amostra é mais reduzida do que seria de esperar, pois tivemos de separar os cursos cujos programas têm características diferentes. Por exemplo, para alguns cursos de Doutoramento o número total de ECTS do curso corresponde a 180, para outros a 240 ECTS. Tendo em conta que, atualmente, a maioria dos cursos desta Universidade são cursos com 180 ECTS, apenas considerámos estas observações. Também no caso dos Doutoramentos, não foi possível obter a informação de todas as datas de aprovação do projeto de dissertação, o que veio reduzir ainda mais a dimensão da nossa amostra.

Verificámos que apenas 21,5% dos alunos desta amostra entregou a sua dissertação de Doutoramento (eventos) e que os restantes alunos estavam ativos ou inativos à data de recolha dos dados (censuras) (Figura 3.7). Também concluímos que dos 158 alunos com ingresso em cursos de Doutoramento, 55,1% são do sexo feminino e 79,1% são de nacionalidade Portuguesa. A média da idade de ingresso destes alunos é igual a 40,1 anos e o desvio padrão igual a 10 anos (Figura 3.8), sendo que 50% dos alunos ingressou na Universidade até aos 38 anos. A maioria destes alunos ingressou na Universidade em 2009 (64,6%), 15,2% ingressou em 2008 e 20,3% em 2007. O curso da maioria dos alunos pertence à escola da Ciências Sociais (60,8%), 25,9% à Escola de Ciências e Tecnologia e apenas 13,3% à Escola de Artes. A média percentagem de ECTS creditados é igual a 14,1%, o desvio padrão igual a 25,2% e metade destes alunos não obtiveram qualquer creditação de ECTS aquando do ingresso na Universidade (Figura 3.8).



**Figura 3.7: Distribuição dos tempos até à entrega da dissertação de Doutoramento (eventos); Distribuição dos tempos até à data de inativação ou de recolha dos dados (censuras).**



**Figura 3.8: Distribuição da idade de ingresso (158 alunos); Distribuição da percentagem de ECTS creditados (158 alunos).**

A duração da parte curricular foi calculada através do tempo que decorre entre a data de ingresso na Universidade e a data da última avaliação. Contudo, no caso dos Doutoramentos, nem todos os cursos têm a mesma duração da parte curricular e alguns até não têm parte curricular. Por isso, optámos por considerar a variável da duração da parte curricular com três categorias: 1 – terminou a parte curricular no tempo previsto (23,9%); 2 – não terminou a parte curricular no tempo previsto (65,9%); 3 – curso sem parte curricular (10,2%).

As variáveis nível de escolaridade dos pais, distrito e tipologia da área de residência e a duração da parte curricular apresentaram alguns valores omissos. Deste modo, apenas pudemos recorrer a 155, 125 e 138 observações, respetivamente. Podemos concluir que, para a maioria dos alunos, ambos os pais têm um nível de escolaridade até ao Secundário (67,7%). Ao contrário do que se verificou nos cursos de Mestrado, o distrito da área de residência da maioria dos alunos de Doutoramento é o Centro (55,2%), sendo que 33,6% provêm do Alentejo e os restantes do Norte, Açores e Algarve. Quanto à tipologia da área de residência, a maioria destes 125 alunos provém de APU (76,8%), 12% de AMU e 11,2% de APR.

Para as restantes variáveis, como a fase de candidatura, número de reingressos, número de anos que os alunos estiveram sem estudar e estatuto de trabalhador estudante, o número de valores omissos era excessivamente elevado. Tendo em conta que já estávamos perante uma amostra muito reduzida, decidimos não considerar estas variáveis para o ajustamento do modelo de regressão de Cox. Podemos desde já adiantar que o facto de termos uma amostra muito reduzida e de não termos acesso a variáveis semelhantes às que Jones-White (2013), Adebi (2987) ou Siegfried (1999) utilizaram no seu estudo, foi de alguma forma condicionante. Efetivamente, não conseguimos ajustar um modelo que representasse minimamente a realidade dos cursos de Doutoramento desta Universidade. Assim sendo, a análise de sobrevivência dos tempos até à

conclusão de uma dissertação de Doutoramento foi realizada apenas através das estimativas de Kaplan-Meier.

### III.4. Análise de Sobrevidência – Mestrados

Para modelar o tempo até à conclusão de uma dissertação de Mestrado, numa primeira fase, procedeu-se a uma análise univariada, na qual obtivemos as estimativas de KM para a mediana do tempo até à conclusão do Mestrado. Depois, comparámos as curvas de sobrevidência de KM para as diferentes categorias das variáveis em estudo e, por fim, ajustámos o modelo de Cox seguindo os passos sugeridos por Hosmer e Lemeshow (2008).

#### III.4.1. Estimativas de Kaplan-Meier

Começamos por destacar alguns resultados obtidos na análise univariada, recorrendo às estimativas de KM para as variáveis consideradas na Tabela 3.1. Na Tabela 3.3, apresentamos as estimativas dos tempos medianos até à entrega da dissertação das categorias de cada variável e os respetivos intervalos de confiança a 95%.

**Tabela 3.3: Estimativas de Kaplan-Meier (KM) para o tempo mediano (me) dos Mestrados, respetivos intervalos de confiança a 95% (IC<sub>95%</sub>) e valores p do teste de Peto ou Log-Rank, dependendo se as curvas de KM se intersectam ou não. (\* variáveis significativas ao nível de significância de 5%)**

Variáveis	me	IC <sub>95%</sub>	Valor p
Sexo			0,02*
Feminino	574	(511 ; 613)	
Masculino	662	(633 ; 728)	
Nacionalidade			0,34
Portuguesa	613	(589 ; 647)	
Outra	591	(504 ; 666)	
Idade			<0,001*
[20 ; 24)	468	(436 ; 481)	
[24 ; 29)	648	(603 ; 690)	
[29 ; 37)	684	(646 ; 816)	
[37 ; 72]	738	(643 ; 809)	
Nível de escolaridade dos pais			0,43
Ambos até ao ensino secundário	569	(501 ; 609)	
Pelo menos um tem ensino superior	511	(479 ; 643)	
NUTSII			0,21
Alentejo	608	(531 ; 648)	
Centro	481	(332 ; NA)	
Algarve	568	(436 ; 759)	

Norte	646	(608 ; 751)	
Açores	448	(314 ; NA)	
Madeira	662	(496 ; NA)	
Tipologia			0,47
AMU	609	(490 ; 797)	
APR	567	(496 ; 646)	
APU	624	(598 ; 661)	
Ano de ingresso			<0,001*
2007	794	(687 ; 912)	
2008	598	(501 ; 642)	
2009	613	(569 ; 662)	
2010	481	(457 ; 491)	
2011	469	(336 ; 596)	
Fase de candidatura			0,11
1ª Fase	527	(492 ; 603)	
2ª Fase	485	(473 ; 792)	
3ª ou 4ª Fase	643	(585 ; 986)	
Escola			0,001*
Ciências Sociais	633	(603 ; 660)	
Artes	858	(531 ; NA)	
Ciências e Tecnologia	567	(500 ; 632)	
Estatuto de trabalhador estudante			0,01*
Não	603	(551 ; 630)	
Sim	739	(643 ; 914)	
Número de reingressos			0,01*
Zero	609	(569 ; 633)	
Um ou mais	1026	(811 ; NA)	
Percentagem de ECTS creditados			0,55
[0 ; 16)	613	(581 ; 648)	
[16 ; 47,4)	642	(517 ; 690)	
[47,4 ; 100)	625	(526 ; 668)	
Duração da parte curricular			0,001*
Terminou no tempo previsto (até 3 semestres)	580	(518 ; 625)	
Terminou depois do tempo previsto (mais de 3 semestres)	661	(613 ; 704)	

Com este estudo, concluímos que, a partir dos 500 dias, os homens têm uma maior probabilidade de sobreviver ao evento. Assim, após este período, são as mulheres que têm uma maior probabilidade de terminar a sua dissertação mais cedo (Figura 3.9). Enquanto metade das mulheres entrega a sua dissertação ao fim de 574 dias, o tempo mediano dos homens é significativamente superior, 662 dias (Tabela 3.3). Podemos ainda observar que para ambos os sexos o período de maior risco de entrega da dissertação ocorre entre o primeiro e o terceiro ano após a data de aprovação do projeto de dissertação.

Um ano após a aprovação do projeto de dissertação, são os alunos mais jovens que entregam a sua dissertação mais cedo (Figura 3.9). Concluímos que apenas existem diferenças

significativas entre as curvas de sobrevivência dos alunos com idades inferiores aos 24 e os restantes. Para os alunos mais jovens o período de maior risco de entrega da dissertação ocorre depois do primeiro ano e até aos 2,5 anos, enquanto que para os alunos com idades superiores ou iguais aos 24 anos esse período ocorre até, aproximadamente, aos 3 anos. Metade dos alunos que ingressam antes dos 24 anos entregam a sua dissertação antes dos 468 dias, já os que ingressam em idades entre os 24 e 28 anos, os 29 e 36 anos e os 37 e 72 anos têm tempos medianos significativamente superiores, 648, 684 e 738 dias, respetivamente (Tabela 3.3).

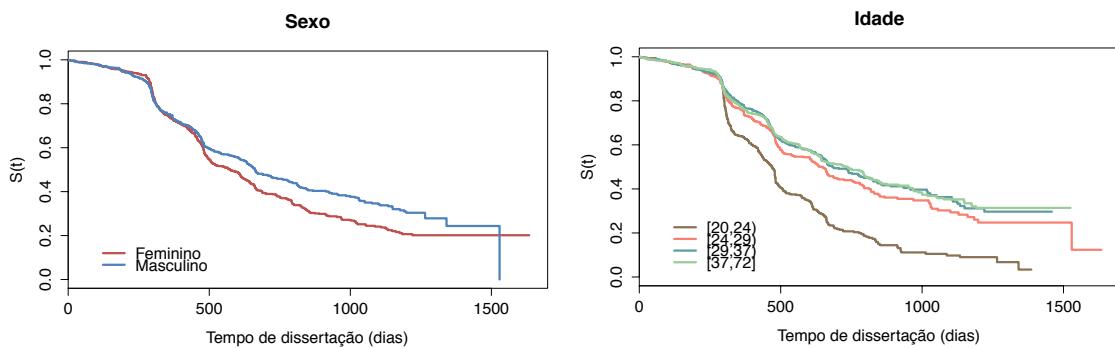


Figura 3.9: Curvas de Kaplan-Meier para o sexo e idade.

São os alunos com ingresso em 2011 que entregam a sua dissertação mais cedo, registando-se um tempo mediano igual a 469 dias. Já os alunos com ingresso em anos anteriores apresentam uma maior probabilidade de prolongar a entrega da sua dissertação, sendo que o período de maior risco ocorre após o primeiro ano (Figura 3.10). Note-se que as curvas de sobrevivência obtidas para os diferentes anos de ingresso diferem significativamente, com a exceção das curvas relativas aos anos de 2008 e 2009. Os tempos medianos registados para os anos de ingresso 2010, 2009, 2008 e 2007 são iguais a 481, 613, 598 e 794 dias, respetivamente (Tabela 3.3).

Após o primeiro ano, os alunos da Escola de Ciências e Tecnologia entregam as suas dissertações mais cedo do que os alunos da Escola de Ciências Sociais, sendo que as restantes curvas da Figura 3.10 não diferem significativamente. O tempo mediano de sobrevivência ao evento é inferior no caso dos alunos que ingressam em cursos da escola Ciências e Tecnologia (567 dias), sendo que este tempo é igual a 633 dias para a Escola de Ciências Sociais e 858 dias para a Escola de Artes (Tabela 3.3).

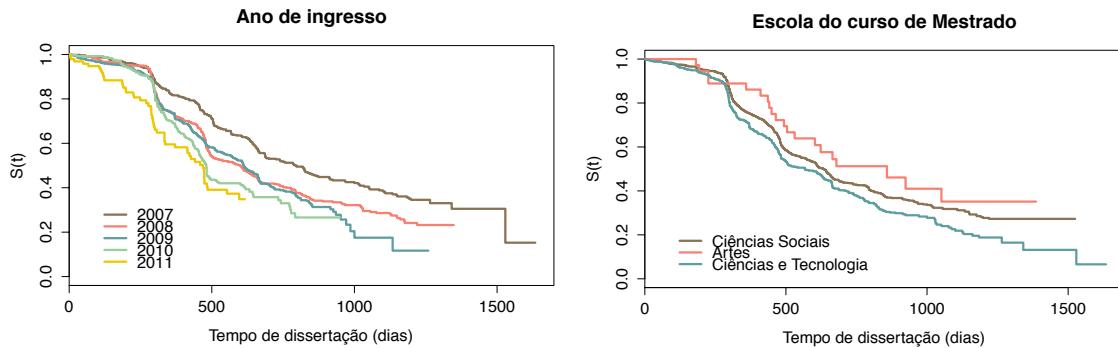


Figura 3.10: Curvas de Kaplan-Meier para o ano de ingresso e escola do curso.

São os alunos que obtiveram o estatuto de trabalhador estudante que apresentam uma maior probabilidade de sobreviver ao evento, sendo que após quase 3 anos esta situação inverte-se (Figura 3.11). O tempo mediano de sobrevivência ao evento é inferior no caso dos alunos que não obtiveram o estatuto de trabalhador estudante (603 dias), sendo que este tempo é igual a 739 dias para aqueles que obtiveram este estatuto (Tabela 3.3). Em ambos os casos, o período de maior risco de entregar a dissertação ocorre depois do primeiro ano e até aos 3 anos.

Até ao primeiro ano, os alunos que reingressam uma ou mais vezes na Universidade entregam as suas dissertações mais cedo (Figura 3.11). Porém, esta situação inverte-se a partir deste período, no qual concluímos que são os alunos que não fizeram qualquer inativação da sua inscrição que entregam as suas dissertações mais cedo. O tempo mediano de sobrevivência ao evento é inferior no caso dos alunos que não efetuaram inativações (609 dias), do que no caso dos que fizeram pelo menos um reingresso na Universidade (1026 dias) (Tabela 3.3). No caso dos alunos que não reingressaram na Universidade, o período de maior risco de entrega da dissertação ocorre a partir do primeiro ano.

Por fim, concluímos que os alunos que terminam a parte curricular no tempo previsto entregam a sua dissertação mais cedo (Figura 3.12). Metade dos alunos que terminam a parte curricular em menos 3 semestres entregam a sua dissertação ao fim de no máximo 580 dias, enquanto o tempo mediano para os alunos que não terminam a parte curricular no tempo previsto é igual a 661 dias (Tabela 3.3). Em ambos os casos, o período de maior risco de entregar a dissertação ocorre depois do primeiro ano e até aos 3,5 anos.

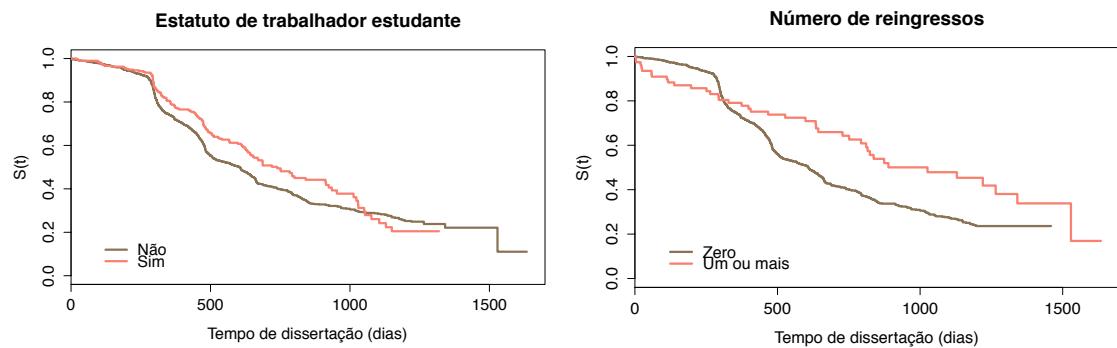


Figura 3.11: Curvas de Kaplan-Meier para o estatuto de trabalhador estudante e número de reingressos.

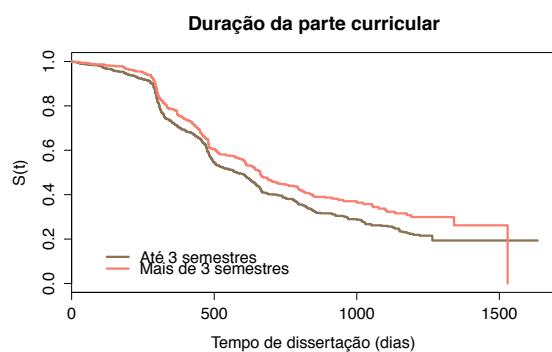


Figura 3.12: Curvas de Kaplan-Meier para a duração da parte curricular.

### III.4.2. Modelo de Cox

De acordo com o primeiro passo sugerido por Hosmer e Lemeshow (2008), para o modelo inicial considerámos as variáveis da Tabela 3.4 cujos valores p do teste de Wald se revelaram inferiores a 0,20.

**Tabela 3.4: Coeficientes estimados ( $\hat{\beta}$ ), respetivos desvios padrão estimados ( $\hat{\sigma}_{\hat{\beta}}$ ) e valores p (teste de Wald) associados, razões de risco ( $\hat{HR}$ ) estimadas e respetivos intervalos de confiança a 95%, para as variáveis que se revelaram significativas no estudo univariado.**

Variáveis	$\hat{\beta}$	$\hat{\sigma}_{\hat{\beta}}$	$\hat{HR}$ $1/\hat{HR}$	$IC_{95\%}(\hat{HR})$ $IC_{95\%}(1/\hat{HR})$	Valor p
<b>Sexo</b>					
Feminino (Referência)					
Masculino	-0,19	0,06	<b>1,21</b>	<b>(1,08 ; 1,36)</b>	0,001
<b>Idade</b>					
	-0,02	0,003	<b>1,02</b>	<b>(1,02 ; 1,03)</b>	<0,001
<b>NUTSII</b>					
Alentejo (Referência)					
Centro	-0,08	0,07	<b>1,08</b>	<b>(0,80 ; 1,06)</b>	0,26
Algarve	0,20	0,15	1,22	(0,91 ; 1,65)	0,19
Norte	-0,01	0,17	<b>1,01</b>	<b>(0,71 ; 1,37)</b>	0,94
Açores	0,37	0,27	1,44	(0,85 ; 2,46)	0,18
Madeira	0,55	0,28	1,74	(1,001 ; 3,01)	0,05
<b>Ano de ingresso</b>					
2007 (Referência)					
2008	0,35	0,08	1,42	(1,21 ; 1,66)	<0,001
2009	0,41	0,08	1,50	(1,27 ; 1,77)	<0,001
2010	0,63	0,09	1,88	(1,57 ; 2,26)	<0,001
2011	0,90	0,15	2,47	(1,82 ; 3,35)	<0,001
<b>Fase de candidatura</b>					
1ª Fase (Referência)					
2ª Fase	-0,04	0,13	<b>1,04</b>	<b>(0,74 ; 1,24)</b>	0,75
3ª ou 4ª Fase	-0,39	0,16	<b>1,48</b>	<b>(1,07 ; 2,05)</b>	0,02
<b>Escola</b>					
Ciências Sociais (Referência)					
Artes	-0,24	0,23	<b>1,28</b>	<b>(0,50 ; 1,22)</b>	0,27
Ciências e Tecnologia	0,19	0,06	1,21	(1,08 ; 1,36)	0,001
<b>Estatuto de trabalhador estudante</b>					
Não (Referência)					
Sim	-0,22	0,09	<b>1,25</b>	<b>(1,05 ; 1,48)</b>	0,012
<b>Número de reingressos</b>					
Zero (Referência)					
Um ou mais	-0,49	0,16	<b>1,63</b>	<b>(1,18 ; 2,25)</b>	0,003
<b>Duração da parte curricular</b>					
Terminou no tempo previsto (até 3 semestres) (Referência)					
Terminou depois do tempo previsto (mais de 3 semestres)	-0,20	0,06	<b>1,22</b>	<b>(1,08 ; 1,37)</b>	0,001

Depois, concluímos que no modelo multivariado apenas as variáveis idade, ano de ingresso, fase de candidatura, estatuto de trabalhador estudante e duração da parte curricular se revelaram significativas a menos de 5%. As variáveis fase de candidatura e duração da parte curricular continham alguns valores omissos, por isso para este modelo apenas pudemos considerar 1429 observações.

Para a variável idade, a única variável contínua do modelo multivariado, o pressuposto da linearidade com o logaritmo da função de risco, não foi verificado. Através do método dos quartis, dos resíduos Martingala e do método dos polinómios fracionários, concluímos que a idade não verifica o pressuposto da linearidade. Desta forma, recorremos ao método dos polinómios fracionários, que sugeriu a transformação:

$$Idade1 = \left( \frac{idade}{10} \right)^{-2} \text{ e } Idade2 = \left( \frac{idade}{10} \right)^{-2} \times \log\left( \frac{idade}{10} \right). \quad (3.14)$$

Para o modelo final obtido (Tabela 3.6), agrupámos a variável ano de ingresso em três categorias. Optámos por agrupar os alunos com ingresso em 2007 e 2008, devido ao número reduzido de dados para o ano de 2007. Repare-se que para amostra inicial o número de alunos com ingresso em 2007 correspondia a 23,6% da amostra (Figura 3.2), contudo o facto de considerarmos a variável fase de candidatura no modelo final fez com não pudéssemos contar com 612 valores omissos que correspondem, maioritariamente, a alunos com ingresso em 2007 (77,8%).

Na fase de modelação também agrupámos as categorias correspondentes aos anos de 2009 e 2010, pois os coeficientes correspondentes do modelo de Cox não eram significativamente diferentes entre si. Esta conclusão foi verificada pelo TRV (valor  $p = 0,91$ ) que comparou este modelo com o modelo com os anos em separado. Também concluímos que os coeficientes do modelo correspondentes à 1<sup>a</sup> e 2<sup>a</sup> fase de candidatura não diferiam entre si (valor de prova do TRV igual a 0,22), por isso agrupámos as categorias correspondentes a estas duas fases de candidatura.

Após o estudo do pressuposto de linearidade das variáveis contínuas e do agrupamento das categorias de algumas variáveis, estudámos as interações que faziam sentido no contexto deste estudo. Concluímos que existem interações significativas entre as variáveis Idade e estatuto de trabalhador estudante e entre as variáveis ano de ingresso e duração da parte curricular.

No ajustamento do modelo de Cox, efetuámos uma análise de resíduos, analisando os resíduos *Deviance* e *Martingala* para detetar *outliers* (Figuras 3.13 e 3.14) e os resíduos *Score* para verificar a influência das observações em cada uma das covariáveis (Figura 3.15). Uma análise a

estes resíduos permitiu identificar a observação 1029 como possível *outlier* e a observação 1374 como influente. Também recorremos a uma análise dos *DfBetas* com o objetivo de identificar observações influentes em cada um dos coeficientes estimados (Figura 3.16) e no conjunto de todos os parâmetros do modelo (Figura 3.17). Esta análise permitiu identificar a observação 1374 que já havia sido identificada através dos resíduos *Score*.

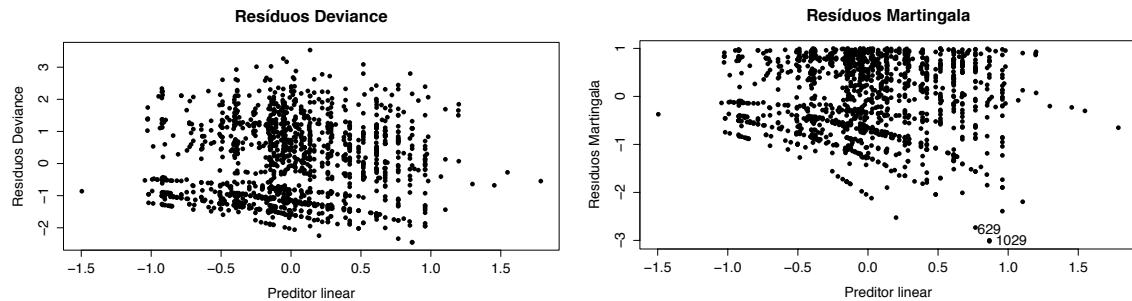


Figura 3.13: Resíduos *Deviance* e *Martingala* para o modelo de Cox ajustado, em função dos valores preditos.

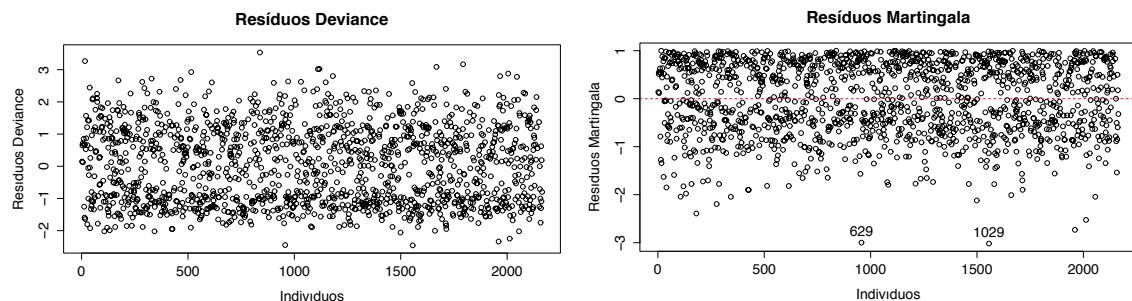


Figura 3.14: Resíduos *Deviance* e *Martingala* para o modelo de Cox ajustado, em função dos indivíduos.

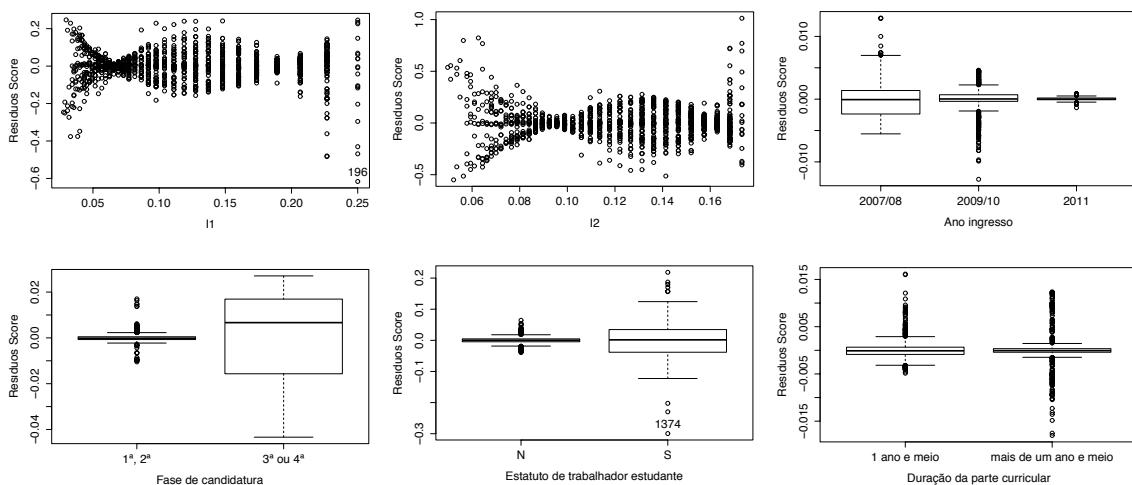


Figura 3.15: Resíduos *Score* para o modelo de Cox ajustado, em função das covariáveis.

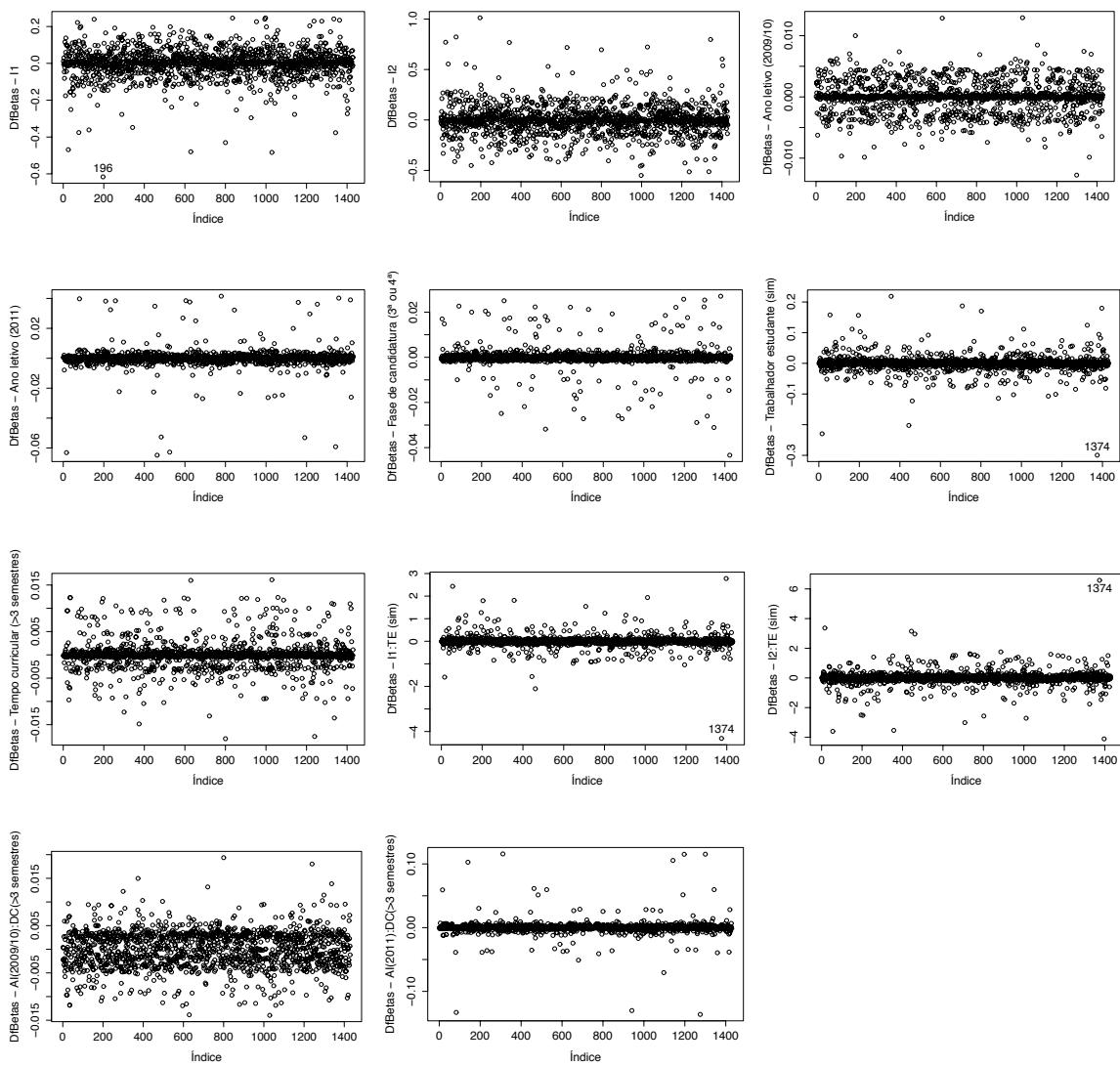


Figura 3.16: DfBetas para cada coeficiente do modelo de Cox ajustado.

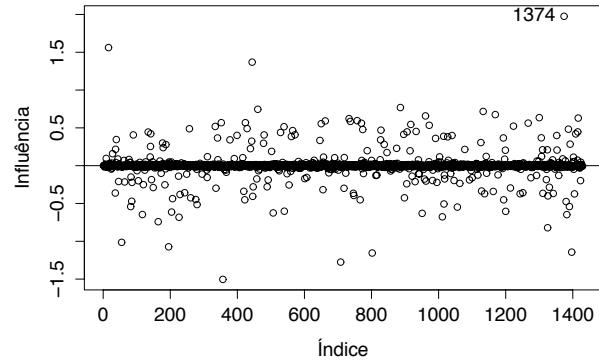


Figura 3.17: DfBetas para o conjunto dos coeficientes do modelo de Cox ajustado.

Tendo em conta as observações identificadas, ajustámos dois novos modelos. No primeiro, retirámos a observação número 1029 e analisámos o seu efeito através da diferença entre os coeficientes estimados do modelo inicial e do modelo sem esta observação. Depois, repetimos este procedimento, mas retirando apenas a observação 1374 ao modelo inicial. Os resultados dos novos modelos, com os novos coeficientes estimados e os respetivos desvios padrão, estão apresentados na Tabela 3.5. No primeiro caso, é possível concluir que a diferença entre a maioria dos coeficientes estimados é relativamente baixa (inferior a 22%) e que a exclusão desta observação pouco melhora a concordância do modelo, sendo que a diferença entre as concordâncias dos modelos com e sem esta observação é inferior a 0,2%. Tendo em conta que as diferenças entre os parâmetros estimados são relativamente baixas, optámos por manter esta observação no modelo inicial. Porém, o efeito da observação 1374 nos coeficientes estimados é consideravelmente superior. No caso dos coeficientes da interação da variável idade e estatuto de trabalhador estudante verificam-se alterações nos coeficientes superiores a 119%. Apesar da diferença entre as concordâncias dos modelos com e sem a observação ser relativamente baixa (0,2%), optámos por retirar esta observação do modelo. O modelo final está apresentado na Tabela 3.6.

**Tabela 3.5: Parâmetros estimados e respetivos desvios padrão estimados do modelo inicial (com todas as observações), do modelo sem a observação 1029 e do modelo sem a observação 1374. Incluímos as diferenças dos coeficientes estimados (em percentagem) entre os novos modelos e o modelo inicial.**

Parâmetros	Modelo com todas as observações <sup>1</sup>		Sem observação 1029			Sem observação 1374		
	$\hat{\beta}$	$\hat{\sigma}_{\hat{\beta}}$	$\hat{\beta}$	$\hat{\sigma}_{\hat{\beta}}$	Dif (%)	$\hat{\beta}$	$\hat{\sigma}_{\hat{\beta}}$	Dif (%)
I1	13,38	2,95	13,88	2,96	-3,7%	13,39	2,95	-0,04%
I2	-14,48	5,52	-15,23	5,52	-5,1%	-14,49	5,52	-0,04%
AI (2009 e 2010)	0,09	0,09	0,08	0,09	14,3%	0,1	0,09	-4,9%
AI (2011)	0,78	0,22	0,77	0,22	0,8%	0,78	0,22	-0,4%
FC (3 <sup>a</sup> ou 4 <sup>a</sup> Fase)	-0,58	0,17	-0,58	0,17	-0,4%	-0,58	0,17	0,2%
TE (Sim)	0,68	1,01	0,66	1,01	3,9%	1,06	1,06	-55,4%
DC (> 3 semestres)	-0,77	0,13	-0,79	0,13	-2,1%	-0,77	0,13	0,2%
I1 x TE	-4,54	10,85	-5,02	10,86	-10,6%	0,88	11,73	119,4%
I2 x TE	-3,29	18,5	-2,59	18,51	21,2%	-11,59	19,73	-252,8%
AI (2009 e 2010) x DC	0,43	0,16	0,44	0,16	-3,4%	0,42	0,16	1,3%
AI (2011) x DC	0,67	0,41	0,68	0,41	-0,9%	0,67	0,41	0,4%

<sup>1</sup> Os valores p do teste de Wald obtidos para o modelo inicial com todas as observações foram, por ordem dos coeficientes da tabela: <0,001; 0,009; 0,32; <0,001; <0,001; 0,50; <0,001; 0,68; 0,86; 0,006; 0,10 (coeficiente de concordância = 0,62;  $R^2$  = 0,11).

**Tabela 3.6: Coeficientes estimados  $(\hat{\beta})$ , respetivos desvios padrão estimados  $(\hat{\sigma}_{\hat{\beta}})$  e valores p (teste de Wald) associados, para o modelo de Cox final (coeficiente de concordância = 0,62;  $R^2 = 0,11$ ).**

Covariável	$\hat{\beta}$	$\hat{\sigma}_{\hat{\beta}}$	Valor p
Idade1 (I1)	13,39	2,95	<0,001
Idade2 (I2)	-14,49	5,52	0,009
Ano de Ingresso (AI)			
2007 e 2008 (referência)			
2009 e 2010	0,1	0,09	0,301
2011	0,78	0,22	<0,001
Fase de Candidatura (FC)			
1 <sup>a</sup> ou 2 <sup>a</sup> (referência)			
3 <sup>a</sup> ou 4 <sup>a</sup>	-0,58	0,17	<0,001
Trabalhador Estudante (TE)			
Não (referência)			
Sim	1,06	1,06	0,315
Duração da parte Curricular (DC)			
≤ 3 semestres (referência)			
> 3 semestres	-0,77	0,13	<0,001
I1 x TE (sim)	0,88	11,73	0,940
I2 x TE (sim)	-11,59	19,73	0,557
AI (2009 e 2010) x DC (> 3 semestres)	0,42	0,16	0,007
AI (2011) x DC (> 3 semestres)	0,67	0,41	0,102

Para o modelo final da Tabela 3.6, validámos o pressuposto de riscos proporcionais através do teste de Harrel e da análise aos resíduos Schoenfeld (Figura 3.18). Os valores p do teste de Harrel associados a todas as variáveis do modelo final foram superiores a 0,05, sendo que obtiveram-se os seguintes valores p, por ordem da Tabela 3.6: 0,16; 0,49; 0,41; 0,32; 0,27; 0,06; 0,24; 0,28; 0,15; 0,34; 0,31. A análise aos resíduos de Schoenfeld, permitiu concluir que todas as variáveis do modelo final apresentaram o comportamento esperado sob o pressuposto dos riscos proporcionais.

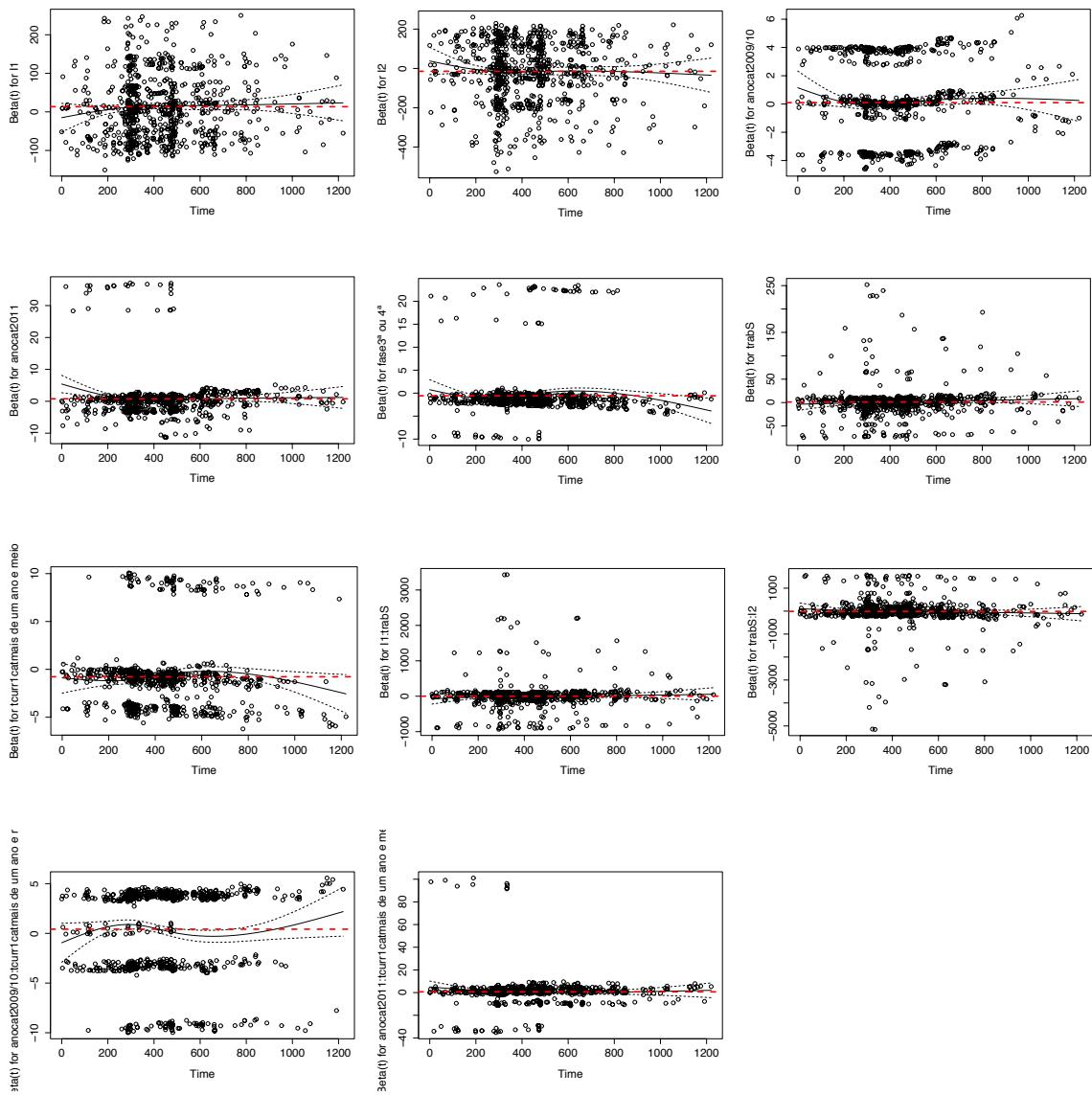


Figura 3.18: Resíduos de Schoenfeld.

O índice de prognóstico associado a este modelo está representado na Figura 3.19, traduzindo-se num modelo aceitável. As linhas a cheio referem-se à função de sobrevida empírica e as linhas a tracejado às estimativas de Kaplan-Meier, obtidas para grupos de alto, médio e baixo índice de prognóstico.

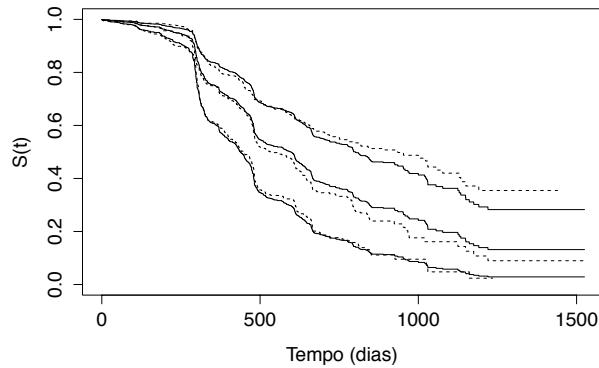


Figura 3.19: Índice de prognóstico do modelo.

### III.4.3. Resultados

As estimativas de Kaplan-Meier para o tempo até à entrega de uma dissertação de Mestrado, permitiram concluir que o período de maior risco de entrega da dissertação ocorre entre o primeiro e o terceiro ano após a data de aprovação do projeto de dissertação, seja qual for a característica do aluno considerada. Os resultados obtidos indicam que as mulheres têm uma maior probabilidade de terminar a sua dissertação mais cedo do que os homens. Adicionalmente, concluímos que os mais jovens, que ingressaram na Universidade nos últimos anos letivos, em cursos da escola de Ciências e Tecnologia, que não obtiveram o estatuto de trabalhador estudante, que não realizaram qualquer inativação da sua inscrição e que terminaram a parte curricular no tempo previsto têm uma maior probabilidade de terminar o curso de Mestrado mais cedo.

A função de sobrevivência estimada (Figura 3.20), obtida a partir do modelo de regressão de Cox da Tabela 3.6, também permitiu concluir que o período de maior risco de terminar a dissertação mais cedo ocorre entre o primeiro e terceiro ano.

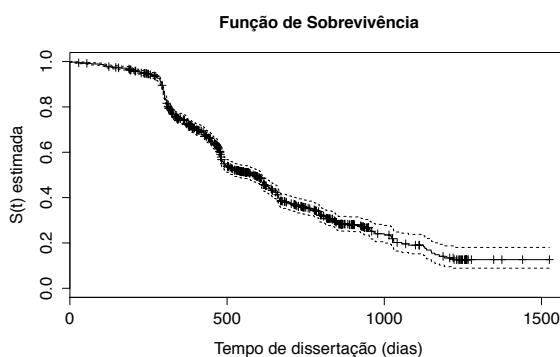
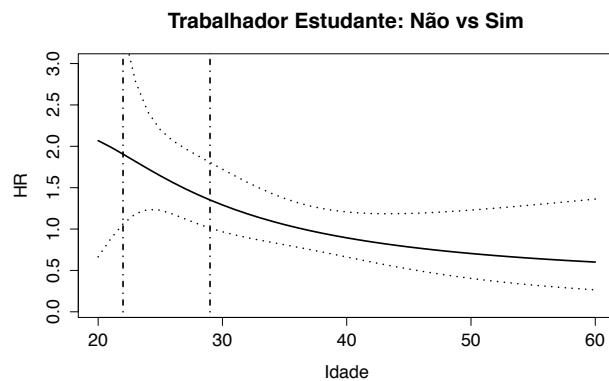


Figura 3.20: Sobrevivência global estimada pelo modelo de Cox ajustado (Mestrados).

A partir do modelo de Cox ajustado, podemos identificar alguns fatores com influência direta na duração da conclusão de um curso de Mestrado. Para valores fixos das restantes variáveis podemos concluir que:

- Um aluno que ingressou na Universidade na primeira ou na segunda fase de candidatura tem um risco quase duas vezes superior ( $1/\text{HR} = \exp(0,58) = 1,78$ ;  $\text{IC}_{95\%} = (1,27; 2,49)$ ) de concluir a dissertação mais cedo.
- Dos alunos com idades compreendidas entre os 22 e 29 anos, os que não obtiveram o estatuto de trabalhador estudante têm um risco maior de terminar a dissertação mais cedo (Figura 3.21). Por exemplo, para os alunos que ingressaram na Universidade aos 22 anos, os que não obtiveram este estatuto têm um risco quase duas vezes superior ( $1/\text{HR} = \exp(-1,06 - 0,88n_1 + 11,59n_2) = 1,91$ ;  $\text{IC}_{95\%} = (1,06; 3,42)$ ) de terminar a dissertação mais cedo do que os que obtiveram este estatuto.<sup>6</sup>



**Figura 3.21: Razão de riscos (HR) para alunos que não obtiveram o estatuto de trabalhador estudante relativamente aos que beneficiaram deste estatuto, para diferentes valores da idade. As linhas verticais marcam as idades entre as quais a diferença é significativa.**

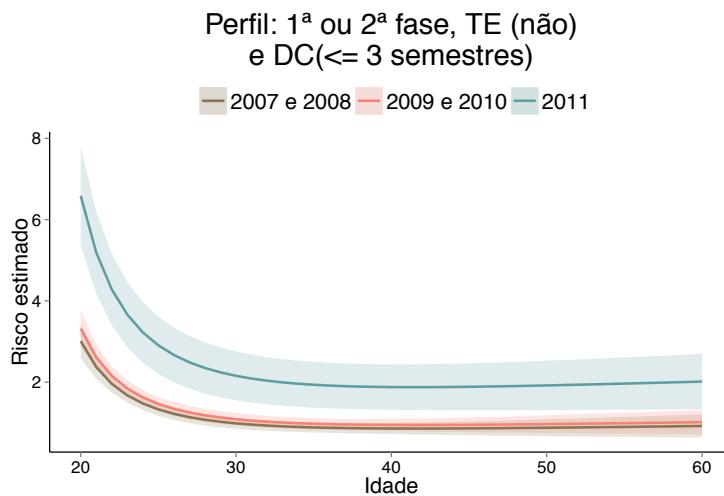
- Dos alunos que demoraram mais de 3 semestres a terminar a parte curricular do seu curso, os que ingressaram em 2009 ou 2010 têm um risco quase duas vezes superior ( $\text{HR} = \exp(0,1 + 0,42) = 1,68$ ;  $\text{IC}_{95\%} = (1,31; 2,15)$ ) de concluir a dissertação mais cedo do que os que ingressaram até 2008. Este risco aumenta para mais de 4 vezes ( $\text{HR} = \exp(0,78 + 0,67) = 4,28$ ;  $\text{IC}_{95\%} = (2,14; 8,56)$ ) para os que ingressaram em 2011 relativamente aos que ingressaram até 2008.
- Dos alunos que terminaram a parte curricular em menos de 3 semestres, apenas se registaram diferenças significativas entre os que ingressaram na Universidade até 2008 e

<sup>6</sup>  $n_1$  e  $n_2$  correspondem à transformação da idade (22 anos), segundo a expressão (3.14), tal que:  $n_1 = (22/10)^{-2}$  e  $n_2 = (22/10)^{-2} \log(22/10)$ .

os que o fizeram em 2011. Concluímos que os alunos com ingresso em 2011 têm um risco quase duas vezes superior ( $HR = \exp(0,78) = 2,19$ ;  $IC_{95\%} = (1,43; 3,35)$ ) de terminar a dissertação mais cedo do que os com ingresso até 2008.

- Dos alunos com ingresso até 2010, quem concluiu a parte curricular mais cedo tem também maior risco de concluir a dissertação mais cedo. Este risco é mais acentuado para os que ingressaram até 2008 ( $1/HR = \exp(0,77) = 2,16$ ;  $IC_{95\%} = (1,68; 2,77)$ ) do que para os que ingressaram em 2009 ou 2010 ( $1/HR = \exp(0,77 - 0,42) = 1,42$ ;  $IC_{95\%} = (1,18; 1,70)$ ).

A partir do modelo de regressão de Cox ajustado, é possível traçar perfis com risco elevado de terminar uma dissertação de Mestrado mais cedo. Como exemplo, apresentamos o perfil para o qual se fixam as características potenciadoras de terminar a dissertação mais cedo. Nesta situação, consideramos indivíduos que ingressaram na primeira ou segunda fase de candidatura, que não obtiveram o estatuto de trabalhador estudante e que terminaram a parte curricular do seu curso no tempo previsto, em função da idade e dos anos de ingresso (Figura 3.22). Concluímos que os alunos com este perfil que ingressaram em 2011 têm um maior risco de terminar a dissertação mais cedo do que os que ingressaram em anos anteriores, seja qual for a idade de ingresso considerada. Por outro lado, para este perfil concluímos que não há diferenças significativas entre os alunos que ingressaram na Universidade nos anos de 2007 ou 2008 e os que ingressaram em 2009 ou 2010. Repare-se que, para os alunos com este perfil, o risco de terminar a dissertação mais cedo é superior para os alunos mais jovens.



**Figura 3.22: Evolução do risco dos alunos que terminam a dissertação mais cedo, em função da idade e dos anos de ingresso, para os alunos com ingresso na 1<sup>a</sup> ou 2<sup>a</sup> fase, que não obtiveram o estatuto de trabalhador estudante e que terminaram a parte curricular no tempo previsto.**

### III.5. Análise de Sobrevivência – Doutoramentos

Para modelar o tempo até à conclusão de uma dissertação de Doutoramento procedeu-se a uma análise univariada, na qual obtivemos as estimativas de KM para a mediana do tempo até à conclusão do Doutoramento. Depois, comparámos as curvas de sobrevivência de KM para as diferentes categorias das variáveis em estudo. Para ajustar um modelo de Cox seguindo os passos sugeridos por Hosmer e Lemeshow (2008), porém, não foi possível obter um modelo que representasse a realidade dos cursos de Doutoramento desta Universidade.

#### III.5.1. Estimativas de Kaplan-Meier

Para a análise de sobrevivência não-paramétrica dos cursos de Doutoramento, destacamos os resultados obtidos na análise univariada, recorrendo às estimativas de KM para as variáveis consideradas na Tabela 3.2. Na Tabela 3.7, podem ver-se as estimativas dos tempos medianos até à entrega da dissertação das categorias de cada variável e os respetivos intervalos de confiança a 95%.

**Tabela 3.7: Estimativas de Kaplan-Meier (KM) para o tempo mediano (me) dos Doutoramentos, respetivos intervalos de confiança a 95% (IC<sub>95%</sub>) e valores p do teste de Peto ou Log-Rank, dependendo se as curvas de KM se intersectam ou não. (\* variáveis significativas ao nível de significância de 5%)**

Variáveis	me	IC <sub>95%</sub>	Valor p
Sexo			0,77
Feminino	1777	(1642 ; NA)	
Masculino	1826	(1614 ; NA)	
Nacionalidade			0,01*
Portuguesa	1824	(1777 ; NA)	
Outra	1459	(1418 ; NA)	
Idade			0,80
[22 ; 33)	1459	(1411 ; NA)	
[33 ; 38)	1778	(1777 ; NA)	
[38 ; 47,8)	1824	(1459 ; NA)	
[47,8 ; 70]	NA	(1642 ; NA)	
Nível de escolaridade dos pais			0,80
Ambos até ao ensino secundário	1824	(1642 ; NA)	
Pelo menos um tem ensino superior	1777	(1645 ; NA)	
NUTSII			0,49
Alentejo	NA	(1158 ; NA)	
Restantes	1824	(1777 ; NA)	
Tipologia			0,03*
AMU	1826	(995 ; NA)	

APR	1158	(1030 ; NA)	
APU	NA	(1777 ; NA)	
Ano de ingresso			0,01*
2007	1778	(1642 ; NA)	
2008	NA	(NA ; NA)	
2009	1139	(1030 ; NA)	
Escola			0,45
Ciências Sociais	1778	(1645 ; NA)	
Artes	1777	(1459 ; NA)	
Ciências e Tecnologia	1459	(1411 ; NA)	
Percentagem de ECTS creditados			0,98
[0 ; 20)	1778	(1645 ; NA)	
[20 ; 50)	NA	(NA ; NA)	
[50 ; 100)	NA	(NA ; NA)	
Duração da parte curricular			0,007*
Terminou no tempo previsto ou sem parte curricular	1824	(NA ; NA)	
Terminou depois do tempo previsto	NA	(NA ; NA)	

Com este estudo, concluímos que os alunos de nacionalidade Portuguesa entregam a sua dissertação mais tarde (Figura 3.23), registando-se um tempo mediano da entrega da sua dissertação igual a 1824 dias, enquanto que os alunos de outra nacionalidade apresentam um tempo mediano significativamente inferior (1459 dias) (Tabela 3.7). Para ambas as nacionalidades, o período de maior risco de entregar a dissertação ocorre depois do segundo ano após a aprovação do projeto de dissertação.

São os alunos que residem em APU que entregam a dissertação mais tarde (Figura 3.23). Para os alunos que residem em AMU e APR, verificamos que não existem diferenças entre as suas curvas de sobrevivências. Porém, concluímos que metade dos alunos que residem em APR entregam a sua dissertação ao fim de 1158 dias, enquanto metade dos alunos que residem em AMU fazem-no ao fim de 1826 dias (Tabela 3.7). Para os alunos que residem em APU não é possível estimar o tempo mediano da entrega da dissertação, pois mais de metade destes alunos não entregaram a sua dissertação. Também neste caso, concluímos que para as três tipologias da área de residência o período de maior risco de entregar a dissertação ocorre a partir do segundo ano.

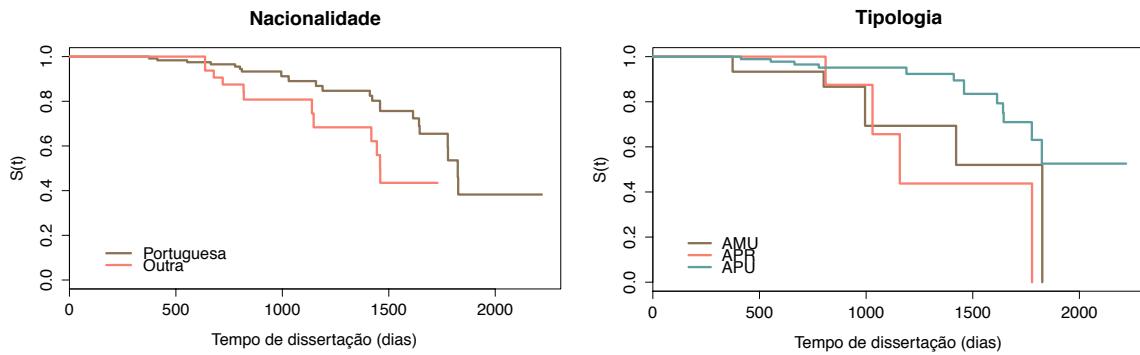


Figura 3.23: Curvas de Kaplan-Meier para a nacionalidade e tipologia da área de residência.

Os alunos com ingresso no ano 2009 entregam a sua dissertação mais cedo do que os alunos com ingresso em anos anteriores (Figura 3.24). Por outro lado, concluímos que não há diferenças significativas entre as curvas de sobrevivência dos anos de ingresso de 2007 e 2008. Para o ano de 2009 regista-se um tempo mediano da entrega da dissertação igual a 1139 dias, enquanto no ano de 2007 esse tempo é consideravelmente superior (1778 dias) (Tabela 3.7). Para qualquer um dos anos considerados o período de maior risco de entrega da dissertação ocorre após o segundo ano.

Por fim, concluímos que os alunos que terminam a parte curricular no tempo previsto, ou aqueles cujos cursos não têm parte curricular, entregam a sua dissertação mais cedo (Figura 3.24), sendo que metade dos alunos que terminam a parte curricular no tempo previsto entregam a sua dissertação ao fim de 1824 dias (Tabela 3.7). Mais uma vez verifica-se que o período de maior risco de entregar a dissertação ocorre a partir do segundo ano, seja qual for a categoria da duração da parte curricular considerada.

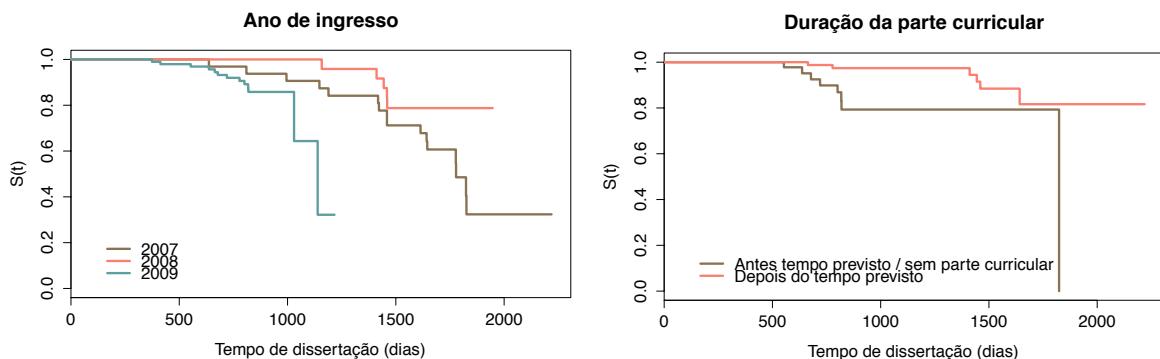


Figura 3.24: Curvas de Kaplan-Meier para o ano de ingresso e duração da parte curricular.

### **III.5.2. Resultados**

As estimativas de Kaplan-Meier para o tempo até à entrega de uma dissertação de Doutoramento, permitiram concluir que o período de maior risco de entrega da dissertação ocorre depois do segundo ano, seja qual for a característica do aluno considerada. Por outro lado, os resultados obtidos indicam que os alunos de outra nacionalidade, residentes em APR ou AMU, com ingresso no ano de 2009 e que terminaram a parte curricular do seu curso no tempo previsto (ou aqueles cujo curso não tem parte curricular) têm uma maior probabilidade de terminar o curso de Doutoramento mais cedo.

### **III.6. Conclusões**

A análise de sobrevivência dos tempos até à entrega de uma dissertação de Mestrado ou de Doutoramento permitiu identificar alguns fatores que influenciam o tempo até à conclusão de um curso. No caso dos Mestrados, este estudo permitiu concluir que o período de maior risco de entrega de uma dissertação ocorre entre o primeiro o terceiro ano. Ao ajustar o modelo de Cox, concluímos que os alunos que ingressaram na Universidade depois da segunda fase de candidatura, nos anos de 2007 ou 2008, que não terminaram a parte curricular no tempo previsto e que beneficiaram do estatuto de trabalhador estudante quando ainda jovens têm um risco muito elevado de demorarem mais tempo a terminar a sua dissertação.

Por outro lado, concluímos que o sexo e a escola do curso não são fatores determinantes no adiamento da conclusão de uma dissertação de Mestrado, ao contrário do que concluíram Teixeira et al. (2012) relativamente aos cursos de 1º Ciclo. Também se conclui que a nacionalidade dos alunos, a tipologia e distrito da área de residência, o nível de escolaridade dos pais, a escola do curso, o número de reingressos e a percentagem de ECTS creditados não são fatores determinantes neste adiamento.

Para os cursos de Doutoramento não foi possível ajustar um modelo de Cox que descrevesse a realidade desta Universidade. Por isso, a análise de sobrevivência dos tempos até à entrega de uma dissertação de Doutoramento apenas contemplou uma análise não paramétrica. Uma das causas responsáveis pode estar relacionada com o facto de termos uma amostra demasiado pequena e de não podermos recorrer a variáveis explicativas que, provavelmente, se adequam melhor a este grau. Efetivamente, alguns dos estudos mencionados referem que variáveis relacionadas com o financiamento dos alunos e o seu grau de satisfação relativamente ao apoio dos

diferentes departamentos e orientadores podem ser preditores do tempo até à conclusão de um curso.

Para os cursos de Doutoramento apenas podemos concluir que há diferenças entre as estimativas de sobrevivência de KM para as categorias das variáveis nacionalidade, tipologia da área de residência, ano de ingresso e duração da parte curricular do curso. Concluímos que os alunos com ingresso em 2007 ou 2008, que são de nacionalidade portuguesa, que residem em APU e não terminam o curso no tempo previsto apresentam uma maior probabilidade de prolongar o tempo de entrega da sua dissertação. Para além disso, seja qual for a categoria da variável considerada o período de maior risco de entregar uma dissertação de Doutoramento ocorre após dois anos da data de aprovação da dissertação. Por outro lado, concluímos que não há diferenças entre as estimativas de sobrevivência das categorias das variáveis sexo, idade, nível de escolaridade dos pais, distrito da área de residência, escola do curso e percentagem de ECTS creditados.



## **Capítulo IV – Perfil dos alunos que pedem semestres adicionais**

### **IV.1. Introdução**

Uma das causas associadas ao prolongamento da conclusão de uma dissertação de Mestrado é o pedido do semestre adicional. Até ao ano letivo de 2013/2014, os alunos poderiam efetuar um pedido de semestre adicional nos Serviços Académicos até ao fim de Setembro do ano limite da entrega da sua dissertação. Depois, poderiam, até Março do ano seguinte, pedir mais um semestre adicional. Efetivamente, os alunos que efetuaram estes pedidos precisaram de mais algum tempo para terminar as suas dissertações. No máximo, quem pediu apenas um semestre adicional prolongou a entrega da dissertação por mais 6 meses. Por outro lado, os alunos que pediram 2 semestres adicionais, precisaram de mais de 6 meses e menos de um ano para entregar a sua dissertação. Assim, para alguns alunos, estes pedidos tiveram um grande impacto no tempo até à conclusão da dissertação.

Cerca de 35% dos alunos desta Universidade que estavam em condições de pedir um semestre adicional para terminar a dissertação de um curso de Mestrado, fizeram-no pelo menos uma vez. Neste capítulo, pretendemos perceber quais os fatores potenciadores ou protetores da realização deste pedido. Desta forma, pretendemos encontrar os fatores que possam ser alvo de políticas de intervenção por parte da Universidade de Évora.

Para a realização deste estudo, ajustámos um modelo de regressão logística com o objetivo de encontrar os determinantes da decisão de pedir pelo menos um semestre adicional e de traçar o perfil mais provável do aluno que efetua este pedido.

### **IV.2. Metodologia aplicada**

Segundo Hosmer e Lemeshow (2013), a aplicação dos modelos de regressão logística tem crescido muito nas últimas duas décadas. Desde a sua aplicação inicial nos estudos epidemiológicos, hoje verifica-se uma crescente aceitação em diversas áreas, tais como na biologia, ecologia, criminologia, engenharia, economia e finanças, entre outras. Esta técnica estatística tem uma grande aplicabilidade quando se pretende ajustar um modelo parcimonioso que permita descrever a relação entre uma variável resposta e um conjunto de variáveis explicativas. Porém, o que distingue um modelo de regressão logística é o facto da variável resposta ser uma variável dicotómica, como é o caso da variável resposta deste estudo (0 – não pediu semestres adicionais; 1

– pediu pelo menos um semestre adicional). Nestes casos, Hosmer e Lemeshow (2013) referem que a aplicação dos modelos de regressão logística tornou-se o método de análise *standard*.

Uma das razões da popularidade deste tipo de modelos está relacionada com o facto da função logística

$$f(z) = \frac{1}{1 + e^{-z}}, \quad (4.1)$$

onde  $z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$ , para  $p$  covariáveis, na qual o modelo de regressão logística se baseia, ser extremamente flexível do ponto de vista matemático e de fácil interpretação. O modelo de regressão logística pode ser escrito através da expressão

$$\pi(\mathbf{X}) = E(Y | \mathbf{X}) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)}}, \quad (4.2)$$

onde  $\pi(\mathbf{X}) = E(Y | \mathbf{X})$  representa a média da variável resposta  $Y$  condicionada aos valores das  $p$  variáveis explicativas, tal que  $\mathbf{X} = (X_1, X_2, \dots, X_p)$ .

Note-se que como a variável resposta é dicotómica, então  $\pi(\mathbf{X})$  representa a probabilidade de ocorrer o evento 1, pelo que apenas toma valores entre 0 e 1. Assim, a transformação de  $\pi(\mathbf{X})$ , através da função *logit*, permite linearizar a expressão 4.2 de tal forma que  $g(\mathbf{X})$

$$g(\mathbf{X}) = \ln\left(\frac{\pi(\mathbf{X})}{1 - \pi(\mathbf{X})}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \quad (4.3)$$

é linear nos seus parâmetros, contínua e varia entre  $-\infty$  a  $+\infty$ . (Hosmer e Lemeshow, 2013)

Os coeficientes do modelo de regressão logística são estimados por máxima verosimilhança, tal que os parâmetros estimados maximizam a função  $L(\mathbf{B})$ ,

$$L(\mathbf{B}) = \sum_{i=1}^n \left\{ y_i \ln[\pi(\mathbf{X}_i)] + (1 - y_i) \ln[1 - \pi(\mathbf{X}_i)] \right\} \quad (4.4)$$

onde:

$\mathbf{B}$  é o vetor  $(\beta_0, \beta_1, \beta_2, \dots, \beta_p)$  dos parâmetros a estimar;

$n$  é o número de observações independentes para o par  $(\mathbf{X}_i, y_i)$ , tal que  $i = 1, 2, \dots, n$ ;

$y_i$  toma o valor da variável resposta (0 ou 1) e  $\mathbf{X}_i$  os valores da variáveis independentes da  $i$ -ésima observação;

$\pi(\mathbf{X}_i)$  é a probabilidade de  $y_i = 1$  condicional a  $\mathbf{X}_i$ ;

$1 - \pi(\mathbf{X}_i)$  é a probabilidade de  $y_i = 0$  condicional a  $\mathbf{X}_i$ .

Para estimar o valor de  $\mathbf{B}$  que maximiza  $L(\mathbf{B})$  é necessário encontrar a solução das  $p+1$  equações de verosimilhança

$$\sum [y_i - \pi(\mathbf{X}_i)] = 0 \quad (4.5)$$

$$\sum_{i=1}^n x_{ij} [y_i - \pi(\mathbf{X}_i)] = 0, \quad j = 1, 2, \dots, p \quad (4.6)$$

Para testar a significância dos coeficientes do modelo comparam-se os valores observados da variável resposta com os valores preditos obtidos a partir dos modelos com e sem a variável explicativa em causa. Esta comparação é baseada no logaritmo da função de verosimilhança definido na equação (4.4), tal que o teste da razão de verosimilhanças pode ser obtido através de:

$$D = -2 \ln \left[ \frac{\text{verosimilhança com a variável } p}{\text{verosimilhança sem a variável } p} \right] \cap \chi^2_{v.g.l.} \quad (4.7)$$

Também é possível testar a significância dos coeficientes através do teste de *Wald*, sendo que este se baseia na comparação da estimativa de máxima verosimilhança do parâmetro  $\hat{\beta}_j$  ( $j = 0, 1, 2, \dots, p$ ) com a estimativa do desvio padrão, sobre a hipótese nula de que  $\hat{\beta}_j = 0$ , tal que:

$$W = \frac{\hat{\beta}_j}{\hat{\sigma}_{\hat{\beta}_j}} \cap N(0,1). \quad (4.8)$$

O intervalo de confiança para o estimador do *logit* pode ser obtido através de:

$$\hat{g}(\mathbf{X}) \pm Z_{1-\alpha/2} V\hat{a}r[\hat{g}(\mathbf{X})]^{1/2}, \quad (4.9)$$

onde  $Z_{1-\alpha/2}$  é o valor obtido a partir da distribuição normal padronizada, tal que a probabilidade do intervalo estimado conter o parâmetro estimado é igual a  $1-\alpha$  ( $\alpha$  é o nível de significância escolhido) e

$$\hat{V\text{ar}}[\hat{g}(\mathbf{X})] = \sum_{j=0}^p X_j^2 \hat{V\text{ar}}(\hat{\beta}_j) + \sum_{k=j+1}^p 2X_j X_k \hat{C\text{ov}}(\hat{\beta}_j, \hat{\beta}_k).$$

A medição do efeito das variáveis explicativas é efetuada estimando a razão de possibilidades (*Odds Ratio*, OR) de um determinado grupo de indivíduos em relação a outro, tal que:

$$\hat{OR} = \frac{\pi(\mathbf{X}^*) / [1 - \pi(\mathbf{X}^*)]}{\pi(\mathbf{X}) / [1 - \pi(\mathbf{X})]}, \quad (4.10)$$

onde  $\mathbf{X} = (X_1, X_2, \dots, X_p)$  e  $\mathbf{X}^* = (X_1^*, X_2^*, \dots, X_p^*)$  denotam os vetores das covariáveis de dois grupos de indivíduos com características distintas.  $X_i^* = 1$  define os indivíduos expostos a determinadas características das variáveis  $i$  e  $X_i^* = 0$  os indivíduos não expostos a essas características.

Por exemplo, se quisermos estimar a razão de possibilidades dos homens em relação à mulheres, então  $\mathbf{X} = (X_1, X_2, \dots, X_p) = X_1$  e  $\mathbf{X}^* = (X_1^*, X_2^*, \dots, X_p^*) = X_1^*$ , onde  $X_1^*$  toma o valor 1 para o grupo dos homens e  $X_1^*$  toma o valor 0 para o grupo da mulheres, tal que:

$$\hat{OR} = \exp[\hat{\beta}_1(X_1^* - X_1)] = \exp[\hat{\beta}_1(1 - 0)] = e^{\hat{\beta}_1} \text{ e } IC_{95\%}(\hat{OR}) = \exp[\hat{\beta}_1 \pm Z_{1-\alpha/2} \sqrt{\hat{V\text{ar}}(\hat{\beta}_1)}].$$

No caso de estarmos perante variáveis contínuas, então a razão de possibilidades pode ser obtida através de:

$$\hat{OR} = \exp[\hat{\beta}_1(X_1 + c) - \hat{\beta}_1(X_1)] = \exp[c\hat{\beta}_1] = e^{c\hat{\beta}_1} \text{ e}$$

$$IC_{95\%}(\hat{OR}) = \exp[c\hat{\beta}_1 \pm Z_{1-\alpha/2} c \sqrt{\hat{V\text{ar}}(\hat{\beta}_1)}].$$

De uma forma geral, o intervalo de confiança associado à razão de possibilidades entre dois grupos (0,1) de uma variável  $X_1$  com interações  $X_1 \times W_1, \dots, X_1 \times W_k$  é calculado pela seguinte expressão:

$$IC_{95\%}(OR) = \exp \left[ \hat{l} \pm Z_{1-\alpha/2} \sqrt{Var(\hat{l})} \right], \quad (4.11)$$

onde  $\hat{l} = \beta_1 + \delta_1 W_1 + \delta_2 W_2 + \dots + \delta_k W_k$ ,

$\beta_1$  é o coeficiente de  $X_1$ ,

$\delta_j$  é o coeficiente de  $X_1 \times W_j, j = 1, \dots, k$  e

$$Var(\hat{l}) = Var(\hat{\beta}_1) + \sum_j W_j^2 Var(\hat{\delta}_j) + 2 \sum_j W_j Cov(\hat{\beta}_1, \hat{\delta}_j) + 2 \sum_j \sum_k W_j W_k Cov(\hat{\delta}_j, \hat{\delta}_k).$$

Para ajustar um modelo multivariado de regressão logística, Hosmer e Lemeshow (2013) sugerem que se efetue os seguintes passos:

1. Para o modelo inicial, selecionar todas as variáveis que se revelam significativas na fase univariada (valores p do teste de Wald inferiores a 0,25);
2. A partir desse modelo eliminar sucessivamente, e por ordem decrescente dos valores p de Wald, todas as variáveis não significativas (valor p > 0,05);
3. Verificar se as variáveis que não foram incluídas no modelo inicial se mostram significativas na presença das restantes, caso em que devem ser adicionadas ao modelo;
4. Ajustar as categorias das variáveis e verificar, para as variáveis contínuas, o pressuposto da linearidade com a função *logit*, através do método dos quartis e pelo método dos polinómios fracionários;
5. Testar, via teste de razão de verosimilhanças, as interações entre as variáveis que fazem sentido no contexto do estudo (valor p ≤ 0,05);
6. Realizar uma análise aos valores atípicos, através dos resíduos de Pearson e Deviance, e às observações influentes nos parâmetros do modelo, através da distância de Cook e dos DFBetas;
7. Avaliar a bondade do ajustamento através do teste de Hosmer e através do coeficiente de determinação de Nagelkerke ( $R_N^2$ ). A estatística do teste de Hosmer é obtida agrupando as observações em decis, de acordo com os valores preditos, e utiliza o

teste do qui-quadrado para comparar os valores observados com os valores preditos em cada grupo.

8. Avaliar a capacidade discriminativa do modelo (capacidade de discriminar entre os indivíduos que experienciaram o evento dos que não o fizeram) através do AUC (Area under the Curve), da sensibilidade (capacidade de detetar o evento corretamente) e da especificidade (capacidade de detetar a não ocorrência do evento corretamente).

### IV.3. Caracterização da amostra

As variáveis explicativas consideradas para o modelo de regressão logístico encontram-se na Tabela 4.1. Note-se que neste caso não pudemos considerar a variável número de reingressos, pois há casos em que o reingresso pode ser efetuado antes do aluno terminar a parte curricular e outros quando o aluno já está a concluir a sua dissertação. Então, se o último reingresso foi efetuado depois do pedido de um semestre adicional não faz sentido considerar esta variável como explicativa para a ocorrência do evento em causa. Também não foi possível recorrer à variável relativa à duração da parte curricular, pois nem todos os alunos tinham terminado a parte curricular aquando do pedido do semestre adicional e não sabemos se terminaram antes ou depois do último pedido de semestre adicional.

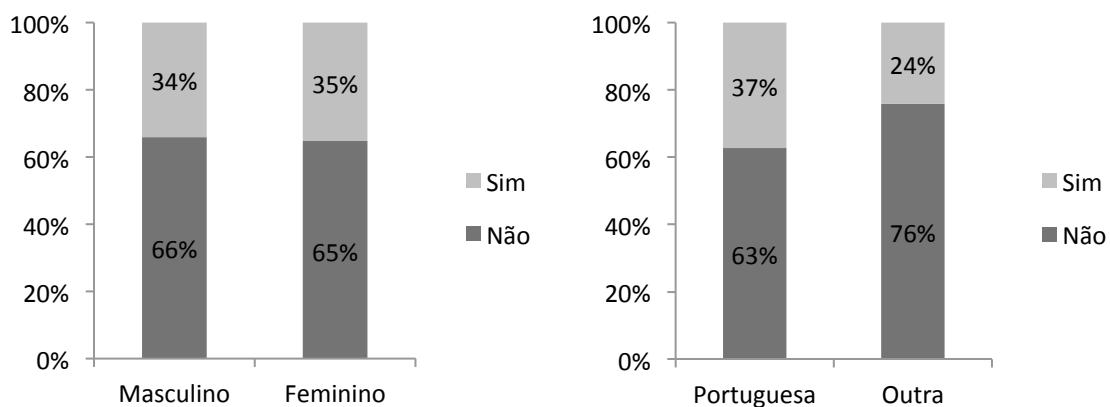
**Tabela 4.1: Variáveis consideradas para o ajustamento do modelo de regressão logística e respetivas categorias.**

Variável	Categoría
Sexo	Feminino
	Masculino
Nacionalidade	Portuguesa
	Outra
Idade	(anos)
Nível de Escolaridade dos pais	Ambos até ao ensino secundário
	Pelo menos um com ensino superior
NUTSII	Alentejo
	Centro
	Algarve
	Norte
	Açores
	Madeira
Tipologia	APU
	AMU
	APR
Ano ingresso (AI)	2007
	2008
	2009

	2010
	2011
Fase de candidatura (FC)	1ª Fase
	2ª Fase
	3ª Fase
Escola do curso	Ciências Sociais (CS)
	Artes (A)
	Ciências e Tecnologia (CT)
Estatuto de trabalhador estudante (TE)	Não
	Sim
Percentagem de ECTS creditados	(%)

Para o estudo do perfil do aluno que pede um ou mais semestres adicionais num curso de Mestrado, considerámos os 1819 alunos diplomados que estavam em condições de fazer este pedido. A maioria destes alunos não pediu qualquer semestre adicional e 35% efetuou este pedido pelo menos uma vez.

Concluímos que 60% dos alunos desta amostra são do sexo feminino e que 81% são de nacionalidade Portuguesa. Também podemos verificar que a proporção de alunos que pediram semestres adicionais é estatisticamente idêntica para os homens e mulheres (valor  $p$  do teste de igualdade de proporções = 0,63), sendo que a maioria destes alunos não pediu semestres adicionais (Figura 4.1). Por outro lado, concluímos que a proporção de alunos que pediu pelo menos um semestre adicional é significativamente superior para os alunos de nacionalidade portuguesa (37%) do que para os alunos de outra nacionalidade (24%) (Figura 4.1), pois o valor de prova obtido no teste de igualdade de proporções é inferior a 0,05.



**Figura 4.1: Distribuição do pedido de semestre adicional pelo sexo; Distribuição do pedido de semestre adicional pela nacionalidade.**

A idade de ingresso destes alunos está entre os 20 e 72 anos, sendo que 50% dos alunos ingressou na Universidade até aos 28 anos (Figura 4.2). A média da idade de ingresso é igual a 30,4 anos e o desvio padrão igual a 8,5 anos. Quando se considera a distribuição da idade pelo pedido de semestre adicional (Figura 4.2), concluímos que não existem diferenças significativas entre as médias das idades de quem não pediu semestres adicionais e de quem o fez pelo menos uma vez (valor de prova do teste de Wilcoxon igual a 0,64).

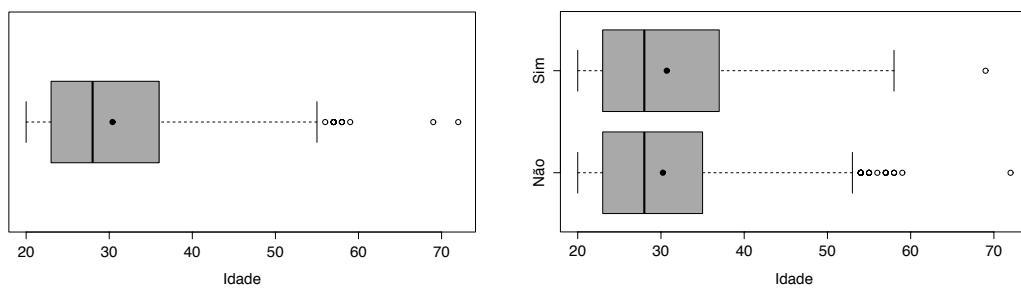


Figura 4.2: Distribuição da idade dos alunos; Distribuição da idade dos alunos por pedido de semestre adicional.

A maioria destes alunos ingressou na Universidade até ao ano de 2009 (79%), sendo que os restantes ingressaram em 2010 ou 2011 (Figura 4.3). Concluímos também que, desde o ano de ingresso de 2008, a percentagem de alunos que pediu semestres adicionais tem vindo a diminuir (Figura 4.3). Dos alunos com ingresso até 2008, 44% efetuou pelo menos um pedido de semestre adicional, enquanto que apenas 4% dos alunos com ingresso em 2011 efetuaram este pedido. Note-se que existem diferenças significativas da proporção de alunos que pediu semestres adicionais entre aqueles que ingressaram até 2008 e em 2009 (valor  $p < 0,001$ ), entre os que ingressaram em 2009 e 2010 (valor  $p = 0,004$ ) e entre os de 2010 e 2011 (valor  $p = 0,002$ ).

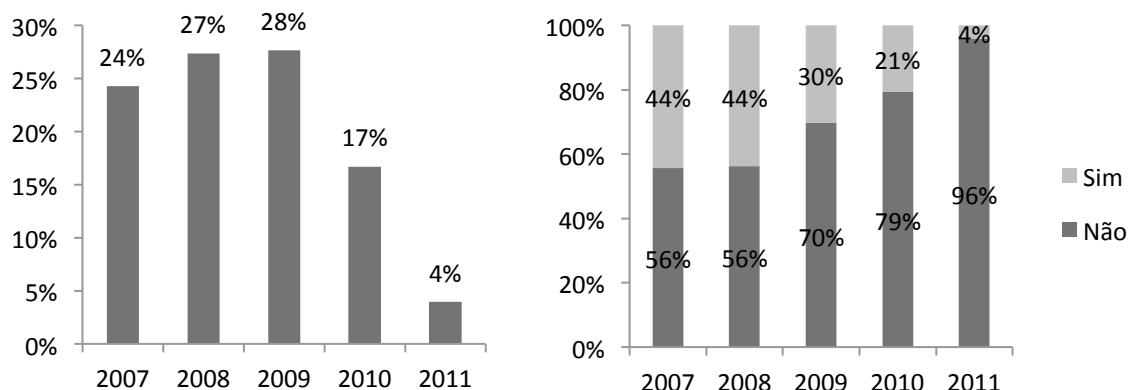
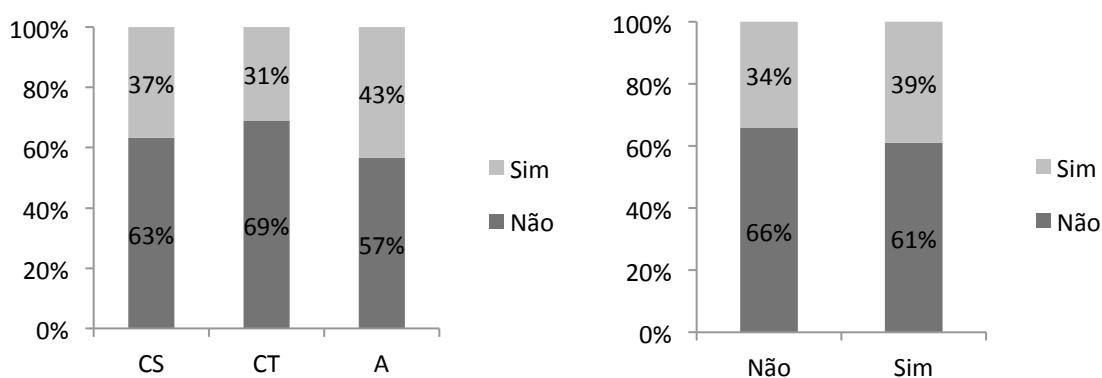


Figura 4.3: Ano de ingresso na Universidade; Distribuição do pedido de semestre adicional pelo ano de ingresso.

Relativamente à escola do curso de Mestrado, concluímos que a maioria dos alunos ingressou em cursos da Escola de Ciências Sociais (57%), 39% da Escola de Ciências e Tecnologia e os restantes da Escola de Artes. Repare-se que, para esta amostra, não se registaram alunos com ingresso em cursos de enfermagem. Para qualquer uma das escolas em causa, concluímos que a maioria dos alunos não pediu semestres adicionais (Figura 4.4). Porém, para os alunos da Escola de Ciências e Tecnologia registou-se a menor proporção de alunos que efetuaram este pedido (31%), sendo que se registam diferenças significativas relativamente à Escola de Artes (valor  $p = 0,03$ ) e à Escola de Ciências Sociais (valor  $p = 0,01$ ) (Figura 4.4).

A esmagadora maioria dos alunos desta amostra não obteve o estatuto de trabalhador estudante (86%). Verificamos que a proporção de alunos que pediu pelo menos um semestre adicional é estatisticamente idêntica (valor  $p = 0,16$ ) para os que obtiveram o estatuto de trabalhador estudante (39%) e para os que não beneficiaram deste estatuto (34%) (Figura 4.4).



**Figura 4.4: Distribuição do pedido de semestre adicional pela Escola do curso; Distribuição do pedido de semestre adicional pelo estatuto de trabalhador estudante.**

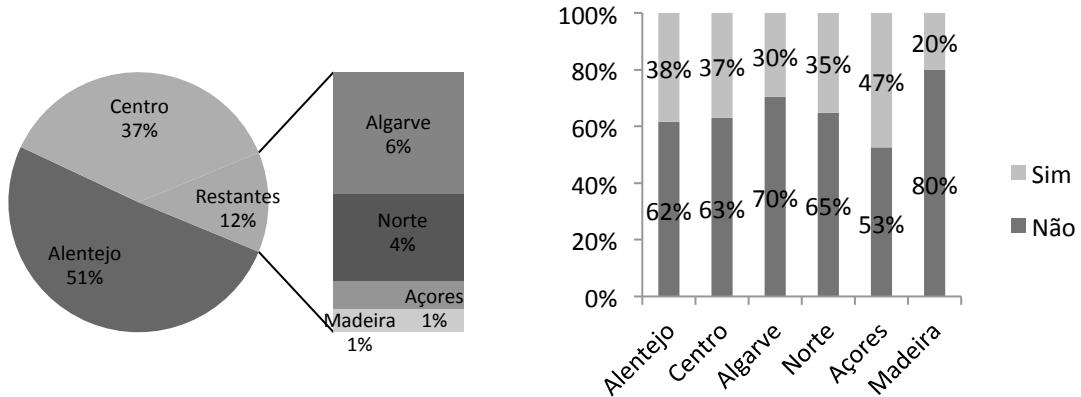
Para as restantes variáveis não foi possível obter o registo de todos os alunos. No caso da variável nível de escolaridade da mãe e do pai apenas pudemos considerar a informação de 1389 alunos. Já no caso do distrito e tipologia da área de residência considerámos 1381 alunos, para a variável fase de candidatura 1209 alunos e para a percentagem de ECTS creditados 1803 alunos. Relativamente à variável número de anos sem estudar, apenas tivemos acesso à informação de 654 alunos, por isso esta variável não foi considerada no ajustamento do modelo de regressão logística.

Para a maioria dos 1389 alunos, ambos os pais têm um nível de escolaridade até ao secundário (77%), sendo que para os restantes alunos pelo menos um dos pais tem um nível de escolaridade superior. Verificamos que a proporção de alunos que pediu pelo menos um semestre

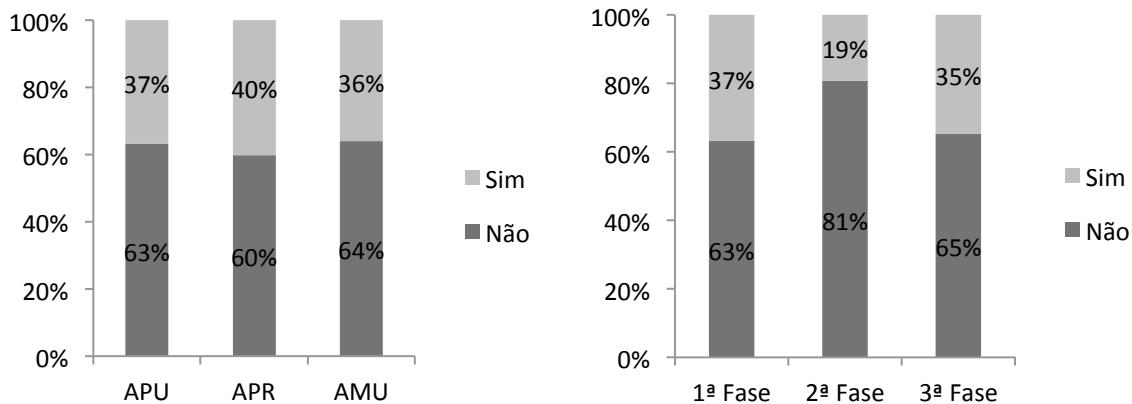
adicional é muito similar para aqueles cujos pais têm até ao ensino secundário (32%) como para aqueles em que pelo menos um dos pais tem o superior (29%) (valor  $p = 0,32$ ).

O distrito da área de residência da maioria dos 1381 alunos é o Alentejo (51%), sendo que é para os alunos que provêm dos Açores que se verifica a maior proporção do pedido de pelo menos um semestre adicional (47%) (Figura 4.5). Contudo, não se registam diferenças significativas entre as proporções dos diferentes distritos de residência (valor  $p = 0,39$ ). Relativamente à tipologia da área de residência, concluímos que a maioria dos alunos provém de APU (76%), 13% de APR e os restantes de AMU. Também neste caso, a proporção de alunos que pediram pelo menos um semestre adicional é estatisticamente idêntica (valor  $p = 0,65$ ) para as diferentes tipologias da área de residência (Figura 4.6).

A esmagadora maioria destes alunos ingressou na Universidade na primeira fase (86%), sendo que 10% ingressou na segunda fase e os restantes na terceira. Ao contrário do que seria de esperar, concluímos que a proporção de alunos que pediram pelo menos um semestre adicional é menor para os alunos que ingressaram na segunda fase (19%) do que para os que ingressaram na primeira ou terceira fase (valor  $p < 0,001$ ).

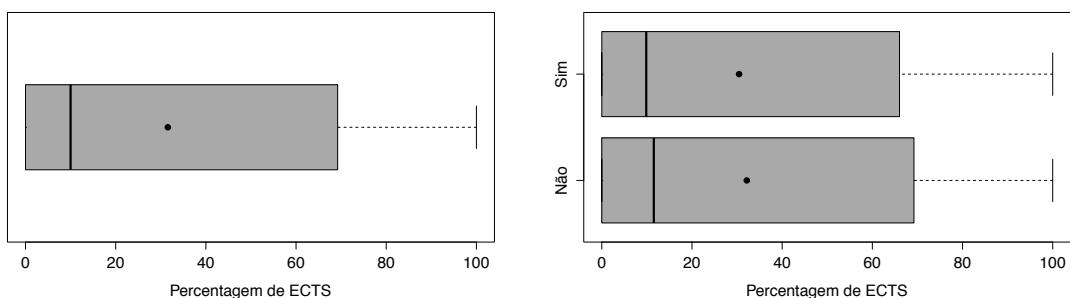


**Figura 4.5: Distrito da área de residência (NUTSII); Distribuição do pedido de semestre adicional por NUTSII.**



**Figura 4.6: Distribuição do pedido de semestre adicional pela tipologia da área de residência; Distribuição do pedido de semestre adicional pela fase de candidatura.**

A percentagem de ECTS creditados aquando do ingresso na Universidade está entre os 0% e 100%, sendo que metade dos alunos obteve até 10% de ECTS creditados (Figura 4.7). A média da percentagem de ECTS é igual a 31,6% e o desvio padrão igual a 37,9%. Quando se considera a distribuição da percentagem de ECTS pelo pedido de semestre adicional (Figura 4.7), concluímos que não existem diferenças significativas entre as médias de quem não pediu semestres adicionais e de quem o fez pelo menos uma vez (valor de prova do teste de Wilcoxon igual a 0,67).



**Figura 4.7: Distribuição da percentagem de ECTS creditados; Distribuição da percentagem de ECTS creditados por pedido de semestre adicional.**

#### IV.4. Modelo de Regressão Logística

De acordo com o primeiro passo sugerido por Hosmer e Lemeshow (2013), para o modelo inicial considerámos as variáveis da Tabela 4.2 cujos valores  $p$  do teste de Wald se revelaram inferiores a 0,25.

**Tabela 4.2: Coeficientes estimados ( $\hat{\beta}$ ), respetivos desvios padrão estimados ( $\hat{\sigma}_{\hat{\beta}}$ ) e valores p (teste de Wald) associados, razões de possibilidades ( $\hat{OR}$ ) estimadas e respetivos intervalos de confiança a 95%, para as variáveis que se revelaram significativas no estudo univariado.**

Variáveis	$\hat{\beta}$	$\hat{\sigma}_{\hat{\beta}}$	$\hat{OR}$ $1/\hat{OR}$	$IC_{95\%}(\hat{OR})$ $IC_{95\%}(1/\hat{OR})$	Valor p
<b>Nacionalidade</b>					
Portuguesa (Referência)					
Outra	-0,62	0,14	<b>1,86</b>	<b>(1,43 ; 2,44)</b>	<0,001
<b>NUTSII</b>					
Alentejo (Referência)					
Centro	-0,06	0,12	<b>1,06</b>	<b>(0,84 ; 1,34)</b>	0,63
Algarve	-0,39	0,26	<b>1,48</b>	<b>(0,90 ; 2,44)</b>	0,13
Norte	-0,14	0,29	<b>1,15</b>	<b>(0,65 ; 2,03)</b>	0,62
Açores	0,37	0,47	1,45	(0,58 ; 3,60)	0,43
Madeira	-0,91	0,65	<b>2,49</b>	<b>(0,70 ; 8,91)</b>	0,16
<b>Ano de ingresso</b>					
2007 (Referência)					
2008	-0,02	0,13	<b>1,02</b>	<b>(0,79 ; 1,32)</b>	0,86
2009	-0,61	0,14	<b>1,84</b>	<b>(1,41 ; 2,40)</b>	<0,001
2010	-1,11	0,17	<b>3,05</b>	<b>(2,18 ; 4,26)</b>	<0,001
2011	-2,91	0,60	<b>18,33</b>	<b>(5,68 ; 59,09)</b>	<0,001
<b>Fase de candidatura</b>					
1ª Fase (Referência)					
2ª Fase	-0,90	0,24	<b>2,46</b>	<b>(1,53 ; 3,94)</b>	<0,001
3ª Fase	-0,09	0,31	<b>1,09</b>	<b>(0,60 ; 2,00)</b>	0,76
<b>Escola</b>					
Ciências Sociais (Referência)					
Artes	0,27	0,23	1,31	(0,83 ; 2,07)	0,25
Ciências e Tecnologia	-0,26	0,10	<b>1,30</b>	<b>(1,06 ; 1,59)</b>	0,01
<b>Estatuto de trabalhador estudante</b>					
Não (Referência)					
Sim	0,21	0,14	1,23	(0,93 ; 1,62)	0,14

Para o modelo multivariado ajustado apenas as variáveis nacionalidade, idade, ano de ingresso, escola do curso e percentagem de ECTS creditados se revelaram significativas a menos de 5%. A variável percentagem de ECTS continha alguns valores omissos, por isso para este modelo apenas pudemos considerar 1803 observações. Repare-se que na fase univariada o valor p do teste de Wald para os coeficientes das variáveis idade e percentagem de ECTS creditados foi superior a 0,25. Contudo, ao testar se as variáveis não incluídas inicialmente no modelo passaram a ser significativas (passo (3) sugerido por Hosmer e Lemeshow (2013)), concluímos que, na presença das restantes variáveis, a idade e a percentagem de ECTS são significativas.

Apenas para a variável idade conseguimos verificar o pressuposto da linearidade com a função *logit*, tanto através do método dos quartis, como através do método dos polinómios fracionários. A variável percentagem de ECTS não verificou este pressuposto e, apesar de termos

considerado a transformação sugerida pelo método dos polinómios fracionários, optámos por categorizar esta variável utilizando como pontos de corte 25% e 50%. Desta forma, no estudo multivariado, recorremos à variável percentagem de ECTS creditados com as seguintes categorias: (0%,25%), (25%,50%) e (50%,100%).

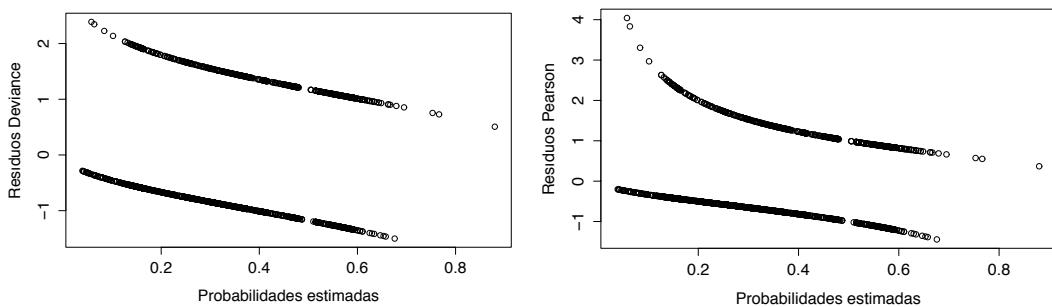
Para o modelo final obtido (Tabela 4.3), agrupámos a variável ano de ingresso em três categorias. Optámos por agrupar os alunos com ingresso em 2010 e 2011, devido ao número reduzido de observações para o ano de 2011. Depois, agrupámos os alunos com ingresso em 2007 e 2008, pois os coeficientes correspondentes do modelo de regressão logística não eram significativamente diferentes entre si. Esta conclusão foi verificada pelo TRV (valor  $p = 0,79$ ) que comparou este modelo com o modelo com os anos em separado. Também concluímos que o coeficiente do modelo correspondente à Escola de Artes não diferia significativamente da referência (Ciências Sociais) (valor de prova do TRV igual a 0,76), por isso agrupámos as categorias correspondentes a estas duas escolas. Por fim, verificámos que o coeficiente do modelo correspondente ao intervalo da percentagem de ECTS (25%,50%) não diferia significativamente do coeficiente relativo ao intervalo (50%,100%) (valor de prova do TRV igual a 0,28), deste modo agrupámos as categorias correspondentes a estes dois intervalos.

Finalmente, estudámos as interações que faziam sentido no contexto deste estudo. Concluímos que existem interações significativas entre a Nacionalidade e a idade dos alunos, entre a idade e a percentagem de ECTS creditados, entre o ano de ingresso e a escola do curso e, finalmente, entre a escola do curso e a percentagem de ECTS creditados. O modelo final obtido está apresentado na Tabela 4.3.

**Tabela 4.3: Coeficientes estimados  $(\hat{\beta})$ , respetivos desvios padrão estimados  $(\hat{\sigma}_{\hat{\beta}})$  e valores p (teste de Wald) associados, para o modelo de regressão logística final.**

Covariável	$\hat{\beta}$	$\hat{\sigma}_{\hat{\beta}}$	Valor p
Constante $(\hat{\beta}_0)$	0,54	0,28	0,06
Nacionalidade			
Portuguesa (referência)			
Outra	-2,66	0,58	<0,001
Idade	-0,01	0,01	0,37
Ano de Ingresso (AI)			
2007 e 2008 (referência)			
2009	-0,88	0,16	<0,001
2010 e 2011	-2,01	0,20	<0,001
Escola do curso			
CS e A (referência)			
CT	-0,46	0,18	0,01
Percentagem de ECTS creditados (ECTS)			
(0%,25%) (referência)			
(25%,100%)	-0,93	0,44	0,03
Nacionalidade (outra) x Idade	0,05	0,02	0,002
Idade x ECTS (25%,100%)	0,03	0,01	0,03
AI (2009) x Escola do curso (CT)	0,63	0,25	0,01
AI (2010 e 2011) x Escola do curso (CT)	1,33	0,32	<0,001
Escola do curso (CT) x ECTS (25%,100%)	-0,72	0,22	0,001

No ajustamento do modelo de regressão logística, efetuámos uma análise de resíduos, através dos resíduos de *Pearson* e *Deviance* para detetar *outliers* (Figura 4.8) e da distância de *Cook* para verificar a influência das observações nos parâmetros estimados pelo modelo (Figura 4.9). A análise à distância de *Cook* permitiu identificar as observações 998 e 361 como influentes. Também recorremos aos resíduos DfBetas para identificar possíveis observações que teriam mais influência nos parâmetros estimados (Figura 4.10), contudo não destacámos mais nenhuma observação.



**Figura 4.8: Resíduos Deviance e Pearson para o modelo de regressão logística ajustado, em função das probabilidades estimadas.**

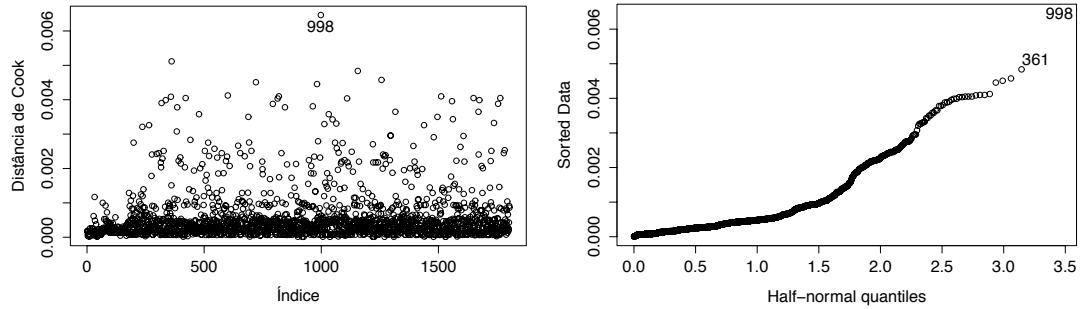


Figura 4.9: Distância de Cook e distância de Cook half-normal para o modelo de regressão logística ajustado, em função dos indivíduos.

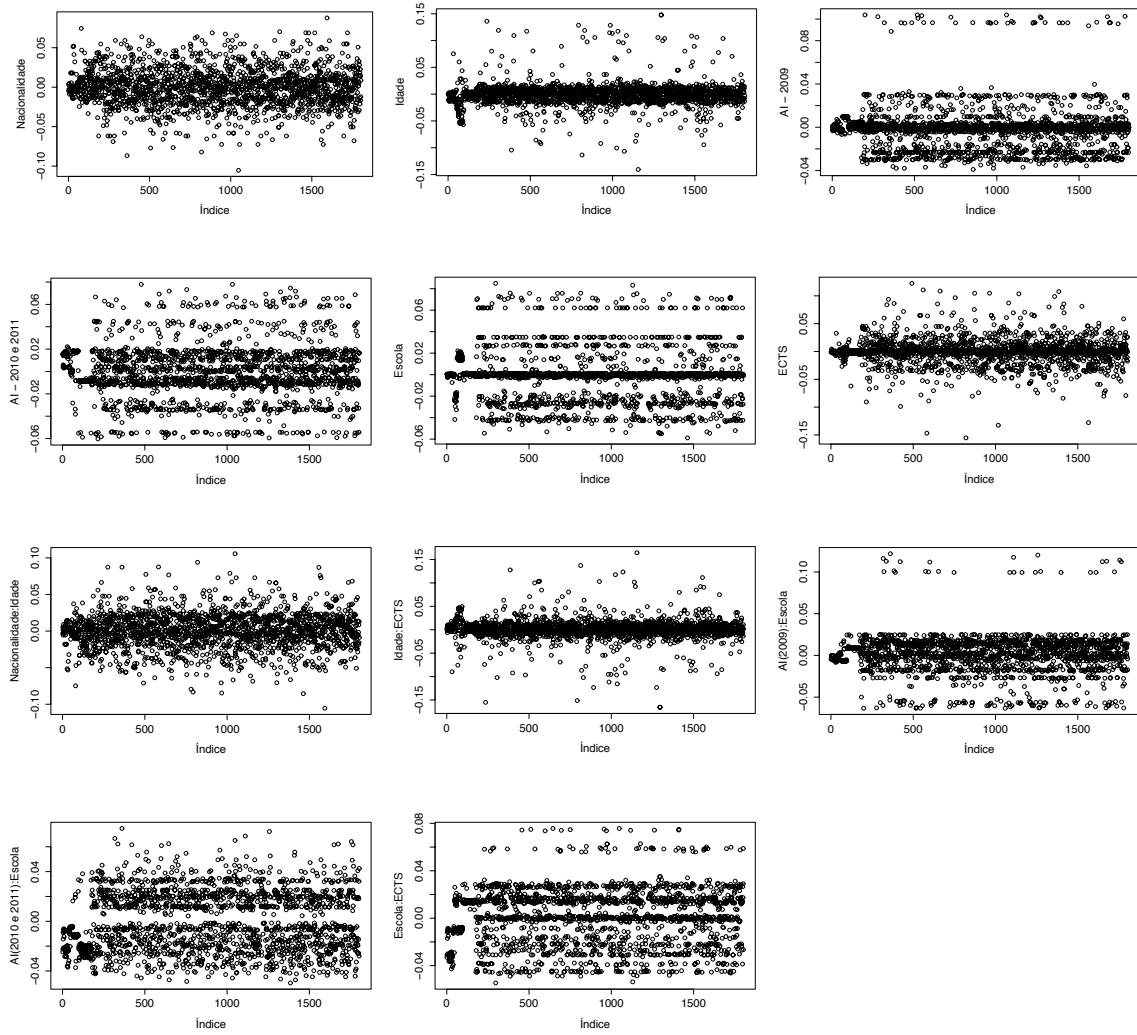


Figura 4.10: DfBetas para cada coeficiente do modelo de regressão logística ajustado, em função dos indivíduos.

Tendo em conta as observações identificadas, ajustámos dois novos modelos. No primeiro, retirámos a observação 998 e analisámos o seu efeito através da diferença entre os coeficientes estimados do modelo inicial e do modelo sem esta observação. Depois, repetimos este procedimento, mas retirando apenas a observação 361 ao modelo inicial. Os resultados dos novos

modelos, com os novos coeficientes estimados e os respetivos desvios padrão, estão apresentados na Tabela 4.4. Em ambos os casos, é possível concluir que a diferença entre a maioria dos coeficientes estimados é relativamente baixa (inferior a 6%), assim como as diferenças entre as *Deviance* dos modelos (inferiores a 0,37%). Deste modo, optámos por não retirar qualquer observação ao modelo da Tabela 4.3.

**Tabela 4.4: Parâmetros estimados e respetivos desvios padrão estimados do modelo inicial (com todas as observações), do modelo sem a observação 988 e do modelo sem a observação 361. Incluímos as diferenças dos coeficientes estimados (em percentagem) entre os novos modelos e o modelo inicial.**

Parâmetros	Modelo com todas as observações		Sem observação 988			Sem observação 361		
	$\hat{\beta}$	$\hat{\sigma}_{\hat{\beta}}$	$\hat{\beta}$	$\hat{\sigma}_{\hat{\beta}}$	Dif (%)	$\hat{\beta}$	$\hat{\sigma}_{\hat{\beta}}$	Dif (%)
Constante ( $\hat{\beta}_0$ )	0,54	0,28	0,55	0,28	-2,89	0,54	0,28	-0,57
Nac. (Outra)	-2,66	0,58	-2,78	0,59	-4,47	-2,68	0,58	-0,67
Idade	-0,01	0,01	-0,01	0,01	-4,00	-0,01	0,01	-1,74
AI (2009)	-0,88	0,16	-0,89	0,16	-0,18	-0,88	0,16	-0,02
AI (2010 e 2011)	-2,01	0,20	-2,02	0,20	-0,25	-2,01	0,20	0,14
Escola (CT)	-0,46	0,18	-0,46	0,18	-1,51	-0,45	0,18	2,07
ECTS (25%,100%)	-0,93	0,44	-0,97	0,44	-4,31	-0,88	0,44	4,82
Nac. (outra) x Idade	0,05	0,02	0,06	0,02	-5,54	0,05	0,02	-1,21
Idade x ECTS (25%,100%)	0,03	0,01	0,03	0,01	-3,83	0,03	0,01	4,65
AI (09) x Escola (CT)	0,63	0,25	0,64	0,25	-1,90	0,63	0,25	0,44
AI (10/11) x Escola (CT)	1,33	0,32	1,34	0,32	-0,86	1,28	0,32	3,47
Escola (CT) x ECTS (25%,100%)	-0,72	0,22	-0,73	0,22	-1,46	-0,74	0,22	-2,73

Em relação à bondade do ajustamento, concluímos que o modelo apresentado na Tabela 4.3 se ajusta bem aos dados, pois para a estatística do teste de Hosmer ( $X^2 = 8,6$ ) obtivemos um valor de prova superior a 0,05 (valor  $p = 0,07$ ). O mesmo se verificou para o valor do  $R^2$  de Nagelkerke, sendo que o valor obtido é igual a 0,15. A curva de ROC (Figura 4.11) obtida permitiu concluir que capacidade do modelo discriminar entre os indivíduos que pediram pelo menos um semestre inicial dos que não o fizeram é aceitável ( $AUC = 0,70$ ), sendo que se obteve uma sensibilidade de 70% e uma especificidade de 61% para o ponto de corte de 0,34. Note-se que a sensibilidade dá-nos a probabilidade de detetar corretamente o pedido de pelo menos um semestre adicional, a especificidade a probabilidade de detetar corretamente zero pedidos de semestres adicionais e o ponto de corte é o ponto que maximiza a sensibilidade e especificidade.

Para validar o modelo recorremos ao método *k-fold cross-validation*. A partir deste método, efetuamos uma partição aleatória da amostra em 10 blocos de igual dimensão, deixando cada

décimo como validação. Este processo resultou num erro estimado relativamente baixo (0,31). Finalmente, repetimos o processo 100 vezes pelo método de *bootstrap*, sendo que obtivemos um AUC igual a 0,703, valor aproximado ao AUC do modelo com todas as observações (0,701). Adicionalmente, para todos os modelos obtidos por *bootstrap* obtivemos valores de prova do teste de Hosmer superiores a 0,05.

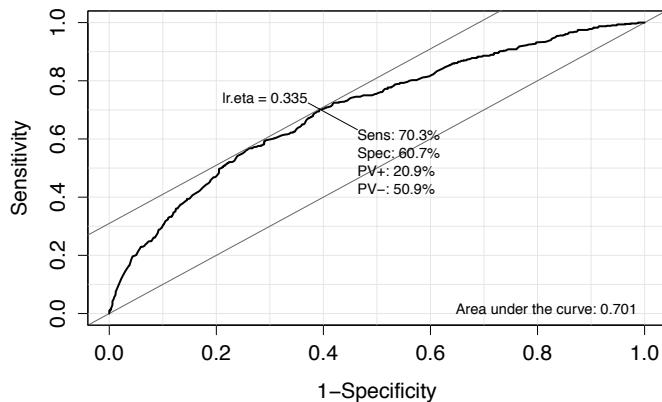


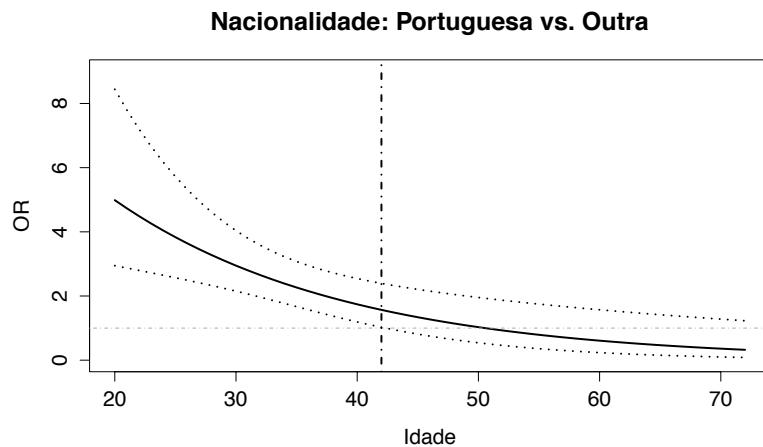
Figura 4.11: Curva de ROC para o modelo de regressão logística ajustado.

## IV.5. Resultados

A partir do modelo de regressão logística ajustado, podemos identificar alguns fatores que potenciam o pedido de um semestre adicional. Para valores fixos das restantes variáveis podemos concluir que:

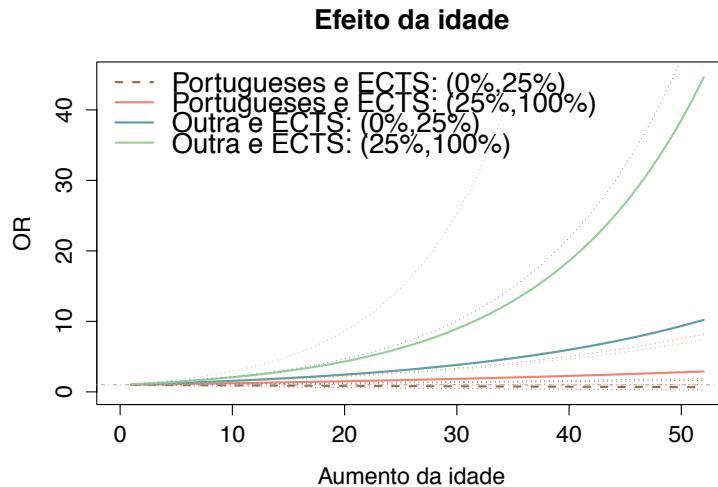
- O efeito da nacionalidade dos alunos na decisão de pedir pelo menos um semestre adicional é influenciado pela idade dos mesmos aquando do ingresso na Universidade de Évora. Concluímos que os alunos de nacionalidade Portuguesa têm mais possibilidades de pedir pelo menos um semestre adicional do que os alunos de outra nacionalidade. Porém, a nacionalidade só é determinante para explicar o facto de um aluno efetuar este pedido quando consideramos os alunos que têm menos de 43 anos (Figura 4.12).

Se considerarmos os alunos que ingressaram na Universidade com 24 anos, concluímos que os que têm nacionalidade Portuguesa têm o quádruplo ( $1/\text{OR} = \exp(2,66 - 0,05 \times 24) = 4,04$ ;  $\text{IC}_{95\%} = (2,65; 6,17)$ ) das possibilidades de pedir pelo menos um semestre adicional. Por outro lado, também concluímos que para os alunos mais velhos o aumento dessas possibilidades atenua-se (Figura 4.12);



**Figura 4.12: Razão de Possibilidades (OR) para alunos que de nacionalidade Portuguesa relativamente aos de outra nacionalidade, para diferentes valores da idade. A linha vertical marca a idade até a qual a diferença é significativa e as linhas a pontilhado são relativas aos intervalos de confiança a 95%.**

- O efeito da idade de ingresso depende a nacionalidade e da percentagem de ECTS creditados. Assim, concluímos que para os alunos que são de outra nacionalidade, por cada ano a mais da idade de ingresso, aumentam as possibilidades destes alunos pedirem semestres adicionais. Contudo, o aumento dessas possibilidades é mais acentuado para os que obtiveram mais do que 25% de ECTS ( $OR = \exp(-0,01 + 0,03 + 0,05) = 1,08$ ;  $IC_{95\%} = (1,04; 1,11)$ ) do que para os que obtiveram entre 0% a 25% ( $OR = \exp(-0,01 + 0,05) = 1,05$ ;  $IC_{95\%} = (1,01; 1,08)$ ). Também concluímos que para maiores diferenças entre as idades dos indivíduos, acentuam-se as possibilidades dos mais velhos pedirem semestres adicionais (Figura 4.13).
- Já para os alunos de nacionalidade Portuguesa, a idade apenas é determinante para aqueles que obtiveram mais do que 25% de ECTS creditados aquando do ingresso na Universidade. Neste caso, concluímos que por cada ano a mais da idade de ingresso, as possibilidades dos mais velhos pedirem semestres adicionais aumentam cerca de 2% ( $OR = \exp(-0,01 + 0,03) = 1,02$ ;  $IC_{95\%} = (1,00; 1,04)$ ).

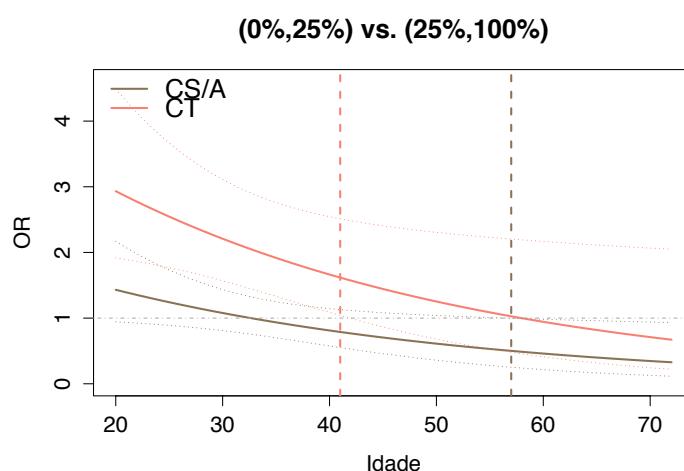


**Figura 4.13:** Razão de Possibilidades (OR) para diferentes aumentos da idade, considerando as diferentes nacionalidades e percentagens de ECTS creditados. As linhas a pontilhado são relativas aos intervalos de confiança a 95%. Note-se que os OR dos indivíduos de nacionalidade Portuguesa que obtiveram entre 0% a 25% de ECTS creditados (linha a tracejado), nunca são significativos seja qual for o aumento da idade considerado.

- O efeito do ano de ingresso na Universidade é determinado pela escola do curso de Mestrado. Dos alunos que ingressaram em cursos da Escola de CS ou A, os que ingressaram nos anos de 2007 ou 2008 têm mais possibilidades de pedir semestres adicionais. Porém, o aumento dessas possibilidades é mais acentuado quando se comparam estes alunos com os que ingressaram depois de 2009 ( $1/OR = \exp(2,01) = 7,47$ ;  $IC_{95\%} = (5,06; 11,03)$ ) do que com os que ingressaram em 2009 ( $1/OR = \exp(0,88) = 2,42$ ;  $IC_{95\%} = (1,78; 3,29)$ ). Deste modo, é possível concluir que, para alunos destas escolas, quanto mais recente é o ano de ingresso, menores são as possibilidades destes alunos pedirem pelo menos um semestre adicional.
- Já quando se consideram os alunos da Escola de CT, apenas se verificam diferenças significativas entre os alunos que ingressaram antes de 2009 e os que o fizeram depois desse ano. Ou seja, concluímos que alunos que ingressaram na Universidade nos anos de 2007 ou 2008 têm o dobro ( $1/OR = \exp(2,01 - 1,33) = 1,98$ ;  $IC_{95\%} = (1,21; 3,25)$ ) das possibilidades de pedir pelo menos um semestre adicional do que os que ingressaram em 2010 ou 2011.
- O efeito da escola do curso é influenciado pelo ano de ingresso e pela percentagem de ECTS creditados. Concluímos que dos alunos que ingressaram na Universidade em 2007 ou 2008, os da Escola de CS ou A têm mais possibilidades de pedir pelo menos um semestre adicional do que os alunos da Escola de CT. Porém, o aumento dessas possibilidades é mais significativo para os alunos que obtiveram mais do que 25% de

ECTS ( $1/\text{OR} = \exp(0,46) = 3,23$ ;  $\text{IC}_{95\%} = (2,24; 4,66)$ ) do que para os que obtiveram menos creditações ( $1/\text{OR} = \exp(0,46 + 0,72) = 1,58$ ;  $\text{IC}_{95\%} = (1,12; 2,23)$ ).

- Já para os alunos que ingressaram em 2009, a escola do curso apenas é determinante para os que obtiveram mais do que 25% de ECTS, sendo que os da Escola de CS ou A têm mais 72% ( $1/\text{OR} = \exp(0,46 + 0,72 - 0,63) = 1,72$ ;  $\text{IC}_{95\%} = (1,05; 2,81)$ ) de possibilidades de pedir pelo menos um semestre adicional do que os alunos da Escola de CT.
- O efeito da escola do curso inverte-se quando consideramos os alunos com ingresso em 2010 ou 2011. Concluímos que para os alunos com ingresso em 2010 ou 2011 com menos de 25% de ECTS creditados, são os da Escola de CT que têm mais ( $\text{OR} = \exp(-0,46 + 1,33) = 2,39$ ;  $\text{IC}_{95\%} = (1,35; 4,24)$ ) possibilidades de pedir pelo menos um semestre adicional do que os das escolas de CS ou A. Note-se que para os que ingressaram nestes anos e que obtiveram mais do que 25% de ECTS, a escola não é significativa para explicar o fato destes alunos pedirem semestres adicionais.
- O efeito da percentagem de ECTS creditados aquando do ingresso na Universidade depende da idade dos alunos e da escola do curso. Assim, concluímos que para os alunos que ingressaram em cursos da Escola de CS ou A depois dos 56 anos, os que obtiveram mais do que 25% de ECTS têm mais possibilidades de pedir pelo menos um semestre adicional do que os que obtiveram poucas creditações (Figura 4.14). Esta situação inverte-se para os alunos cujos cursos pertencem à Escola de CT, sendo que para estes alunos com menos de 41 anos, são os que obtiveram menos de 25% de ECTS que têm mais possibilidades de pedir pelo menos um semestre adicional (Figura 4.14).



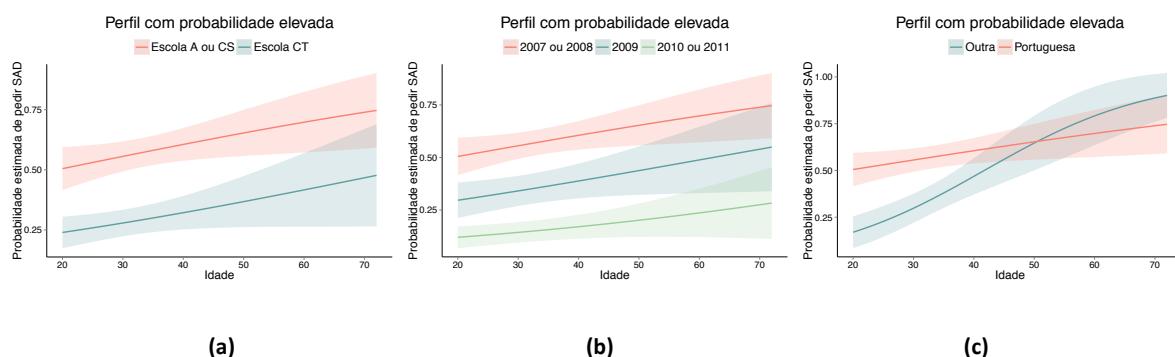
**Figura 4.14:** Razão de Possibilidades (OR) para alunos que obtiveram menos de 25% de ECTS creditados relativamente aos que obtiveram mais do que 25% de ECTS, para diferentes valores da idade e diferentes escolas. As linhas verticais marcam as idades até (CT) ou a partir (CS/A) das quais as diferenças são significativas. As linhas a pontilhado são relativas aos respetivos intervalos de confiança a 95%.

A partir do modelo de regressão logística ajustado, é possível traçar perfis com probabilidade elevada de pedir pelo menos um semestre adicional. Como exemplo, apresentamos o perfil para o qual se fixam as características potenciadoras de efetuar este pedido. Nesta situação, consideramos os alunos que são de nacionalidade Portuguesa, que ingressaram na Universidade de Évora no ano de 2007 ou 2008, na Escola de CS ou A, e que obtiveram mais de 25% de ECTS, em função da idade. Note-se que as probabilidades dos indivíduos com estas características estão representadas nos três gráficos da Figura 4.15 (linhas vermelhas).

Na Figura 4.15 (a), fixamos as características anteriores e fazemos variar tanto a idade como a escola do curso. Repare-se que nestas condições, existem diferenças significativas entre as probabilidades dos alunos da Escola de CS ou A e os da Escola de CT pedirem pelo menos um semestre adicional. Assim, concluímos que são os da Escola de CS ou A que, até aos 60 anos, apresentam sempre maiores probabilidades de pedir pelo menos um semestre adicional.

Já quando se faz variar o ano de ingresso (Figura 4.15 (b)), é possível verificar que quanto mais recente é o ano, menores são as probabilidades destes alunos pedirem semestres adicionais. Concluímos também que, seja qual for a idade dos indivíduos, há sempre diferenças significativas entre quem ingressou antes de 2009 e depois desse ano. Por outro lado, verifica-se que a probabilidade dos alunos que ingressaram em 2009 pedirem pelo menos um semestre adicional só difere dos que ingressaram nos restantes anos quando consideramos alunos que ingressaram até aos 50 anos (2007 ou 2008) ou até aos 58 anos (2010 ou 2011).

Por fim, analisamos o perfil em que fazemos variar a nacionalidade dos alunos (Figura 4.15 (c)). Concluímos que até aos 39 anos, os que têm nacionalidade portuguesa apresentam uma maior probabilidade de efetuarem este pedido. A partir dos 51 anos esta situação parece se inverter, porém não podemos afirmar que existam diferenças significativas entre estas probabilidades.



**Figura 4.15: Perfil de probabilidade elevada de pedir pelo menos um semestre adicional (Portugueses, com ingresso em 2007 ou 2008, na Escola de CS ou A e que obtiveram mais de 25% de ECTS creditados – linhas vermelhas), em função da idade e: da escola do curso (a); do ano de ingresso (b); da nacionalidade dos alunos (c). As regiões a sombreado representam os respetivos intervalos de confiança a 95%.**

#### IV.6. Conclusões

Ao ajustar o modelo de regressão logística foi possível encontrar alguns fatores determinantes para a decisão dos alunos pedirem semestres adicionais. Concluímos que a nacionalidade, a idade, o ano de ingresso, a escola do curso e a percentagem de ECTS creditados são fatores significativos para o explicar o facto destes alunos efetuarem este pedido. Por outro lado, verificamos que o sexo dos alunos, assim como o nível de escolaridade dos pais, o distrito e tipologia da área de residência, a fase de candidatura ou o facto de terem beneficiado do estatuto de trabalhador estudante, não são fatores significativos para explicar o evento em causa.

Este estudo também permitiu traçar o perfil mais provável do aluno que efetua o pedido de pelo menos um semestre adicional. Concluímos que são os alunos de nacionalidade portuguesa, que ingressaram antes de 2009, na Escola de CS ou A e que obtiveram mais do que 25% de ECTS que apresentam uma maior probabilidade de pedir pelo menos um semestre adicional. Também concluímos que quanto maior é a idade de ingresso, maiores são as probabilidades destes alunos efetuarem este pedido.

Se, por um lado, não é possível controlar a idade, o ano de ingresso ou a nacionalidade dos alunos, por outro, é possível averiguar quais as características que distinguem os cursos das escolas de A ou CS dos cursos da Escola de CT. Esta poderá ser uma forma de perceber porque é que os alunos das escolas de A ou CS apresentam maiores probabilidades de efetuar este pedido, seja qual for o cenário considerado.

O facto do efeito da percentagem de ECTS creditados também depender da escola do curso pode suscitar algumas dúvidas. Repare-se que quando consideramos apenas os alunos das escolas de A ou CS, são os que obtiveram mais creditações aquando do ingresso na Universidade que apresentam maiores possibilidades de pedir semestres adicionais. Já quando se consideram os alunos da Escola de CT, são os que obtiveram poucas creditações que apresentam maiores possibilidades de efetuar este pedido.

Por outro lado, podemos verificar uma tendência positiva nos últimos anos, na medida em que concluímos que os alunos que ingressaram na Universidade em anos letivos mais recentes apresentam menores possibilidades de pedir semestres adicionais. Seja qual for o cenário considerado, até mesmo para aqueles alunos que apresentam um perfil mais provável de pedir semestres adicionais, ao longo dos últimos anos verificámos um decréscimo das probabilidades dos alunos desta Universidade efetuarem este pedido.

## **Capítulo V – Controlo de Qualidade dos tempos de procedimentos iniciados pelos alunos**

### **V.1. Introdução**

Os processos de controlo estatístico (SPC – *Statistical Process Control*) visam melhorar a capacidade de um processo através da deteção de alterações que possam surgir ao longo do tempo. Segundo Montgomery (2009), as cartas de controlo estatístico são a forma mais simples de avaliar e controlar o desempenho do processo ao longo do tempo. Uma carta de controlo permite distinguir entre causas de variação inerentes ao processo e causas de variação não naturais, que devem ser detetadas e eliminadas se tal contribuir para melhorar o processo em causa.

As cartas de controlo são uma das principais ferramentas dos processos de controlo estatístico. Estas cartas foram concebidas por Walter A. Shewhart (em 1924), a pensar na monitorização de processos industriais. Porém, têm surgido cada vez mais aplicações em diversas áreas como na economia, finanças, biologia, saúde, entre outras. Em Montgomery (2009), é possível ver-se uma diversidade de aplicações das cartas de controlo de qualidade.

Existem vários tipos de cartas de controlo estatístico, mas, de uma forma geral, as cartas de controlo são uma representação gráfica de valores de uma estatística da característica da qualidade que interessa estudar, em função do tempo, e permitem:

1. Detetar a ocorrência de causas de variação assinaláveis.
2. Controlar o processo ao longo do tempo.
3. Avaliar a capacidade do processo em cumprir determinadas especificações.

Neste capítulo, recorremos a alguns processos de controlo estatístico com o objetivo de identificar causas responsáveis pela variação não natural de alguns processos administrativos desta Universidade. Os processos administrativos aos quais nos referimos, estão associados a alguns dos procedimentos iniciados pelos alunos, nomeadamente ao T proc, T sac, T formação do júri, T reunião, T provas, TD grau e TD curso (Figura 1.1, pág. 4). Efetivamente, referimo-nos a processos que, de uma forma ou outra, podem antecipar ou atrasar ainda mais a conclusão dos cursos de Mestrado.

Para os tempos em análise, elaboramos cartas de controlo de qualidade retrospetivas e medimos índices de capacidade dos serviços. Assim, pretendemos que este estudo ajude a Universidade a implementar ações com vista a reduzir a variabilidade destes processos e a trazer os serviços em causa para valores alvo previamente especificados.

## V.2. Metodologia aplicada

*Total Quality Management (TQM)* é uma estratégia de implementação e gestão da melhoria de qualidade das atividades de uma organização. Segundo Montgomery (2009), este conceito teve início em 1980 com os filósofos Edwards Deming e Joseph Juran. O conceito de *TQM* tem vindo a evoluir ao longo dos anos para um largo espectro de conceitos e ideias envolvendo organizações mais participativas na cultura do trabalho e mais focalizadas nos clientes. Integrando sistemas de qualidade total nas organizações é possível obter e melhorar a qualidade das mesmas. Para tal, é necessário melhorar o desempenho dos processos através de uma abordagem sistemática e estruturada que se desenvolve ao longo de cinco fases, mais conhecidas como *DMAIC: Definir, Medir, Analisar, Melhorar (Improve) e Controlar*. (Montgomery, 2009)

*Definir* o problema passa por selecionar a característica de saída e identificar as variáveis de entrada e saída do processo em questão. *Medir*, significa que se devem definir padrões de desempenho, validar o sistema de medição, estabelecer a capacidade do processo e definir objetivos de desempenho. *Analizar*, permite identificar fontes de variação e selecionar causas potenciais. *Melhorar*, passa por descobrir relações entre variáveis, estabelecer tolerâncias de operação, validar o sistema de medição e determinar a capacidade do processo. *Controlar*, implica implementar controlos do processo.

Para controlar alguns dos processos administrativos desta Universidade implementamos cartas de controlo retrospectivas com o intuito de compreender melhor os processos em causa, para tentar detetar causas responsáveis pela variação dos mesmos e, também, para incentivar a sua utilização. As cartas de controlo implementadas têm como objetivo monitorizar a média dos diferentes tempos em estudo. Contudo, o tipo de cartas de controlo estatístico a implementar depende da distribuição dos tempos dos procedimentos administrativos. Neste sentido, nas secções seguintes apresentamos três cartas distintas, as cartas de controlo de Shewhart, as cartas de controlo EWMA e as cartas de controlo *Skew-Normal*.

Tendo em conta que determinar a capacidade do processo é uma parte importante do processo DMAIC, terminamos esta secção apresentando alguns métodos de investigação da capacidade do processo.

### V.2.1. Cartas de Controlo Shewhart

Uma carta de controlo típica é uma representação gráfica da característica da qualidade que foi medida a partir de uma amostra (de dimensão igual ou superior a 1), em função do tempo ou do número da amostra. Neste estudo, consideramos amostras de dimensão igual a 1 e as características das qualidades a monitorizar são expressas em termos de um valor numérico, mais precisamente, os tempos (dias) até à conclusão de alguns processos administrativos iniciados pelos alunos nos SAC.

As cartas de controlo por variáveis são as mais indicadas para monitorizar a média e a variabilidade dos processos. Estas cartas contêm uma linha central (CL – *Central line*) que representa o valor médio da característica da qualidade correspondente e duas linhas horizontais, chamadas limite superior de controlo (UCL – *Upper control limit*) e limite inferior de controlo (LCL – *Lower control limit*). Estes limites de controlo são escolhidos de modo a que, se o processo está controlado, quase todos os pontos amostrais estarão entre eles. Enquanto os pontos amostrais se mantiverem dentro dos limites de controlo, assumimos que o processo está controlado e que nenhuma ação é necessária. No entanto, um ponto que estiver fora destes limites é interpretado como uma evidência de que o processo está fora de controlo e que é necessário investigar e estabelecer medidas para eliminar a(s) causa(s) responsáveis por tal comportamento. (Montgomery, 2009)

Estas cartas só devem ser utilizadas se a variável em causa tiver uma distribuição normal. Suponhamos que a característica da qualidade que foi medida tem distribuição normal com média  $\mu$  e desvio padrão  $\sigma$ , em que ambos os parâmetros são conhecidos. Se  $x_1, x_2, \dots, x_n$  é uma amostra de dimensão  $n$ , então a média da amostra é dada por

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (5.1)$$

e sabemos que  $\bar{x}$  é normalmente distribuída com média  $\mu$  e desvio padrão  $\sigma_{\bar{x}} = \sigma / \sqrt{n}$ .

Adicionalmente,  $1 - \alpha$  é a probabilidade de que a média de qualquer amostra esteja entre

$$\mu + Z_{\alpha/2} \sigma_{\bar{x}} = \mu + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad \text{e} \quad \mu - Z_{\alpha/2} \sigma_{\bar{x}} = \mu - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}. \quad (5.2)$$

Assim, se  $\mu$  e  $\sigma$  são conhecidos, então as equações (5.2) podem ser usadas como os limites de controlo superiores e inferiores nas cartas de controlo para médias.

É usual substituir  $Z_{\alpha/2}$ <sup>7</sup> por 3 para que se possa implementar os limites de controlo 3-sigma. As cartas de controlo por variáveis, obtidas a partir dos limites de controlo 3-sigma, são chamadas as cartas de controlo de Shewhart (Cartas X para n=1 e Cartas  $\bar{X}$  para n>1). Os limites 3-sigma foram definidos pela primeira vez por Walter A. Shewhart e indicam que a distância dos limites de controlo à linha central, expressa em unidades do desvio padrão, é igual a 3. Ao recorrer ao limites de controlo 3-sigma assumimos a probabilidade de cometer um erro do tipo I (considerar um ponto fora de controlo, quando não existe uma causa assinalável para tal) é igual a 0,27%. Esta probabilidade indica que em média são geradas 370,4 amostras até obtermos um ponto fora de controlo ( $ARL_0$  – Average Run Length). (Montgomery, 2009)

Se a média de uma amostra estiver fora dos limites 3-sigma, então temos uma indicação de que a média do processo deixou de ser igual a  $\mu$ . Na prática, normalmente, não se conhece  $\mu$  e  $\sigma$ , por isso estes parâmetros devem ser estimados a partir de amostras iniciais, normalmente de 20 a 25 amostras, retiradas de um processo sob controlo. Contudo, como estamos perante a implementação de cartas retrospectivas, esta abordagem não será efetuada e implementaremos cartas de controlo considerando os parâmetros  $\mu$  e  $\sigma$  observados para as amostras em causa, mas com o processo sob controlo.

Para verificar se o processo estava controlado para as amostras iniciais, primeiro, é necessário construir a carta de controlo a partir destas amostras. Se todos os pontos estiverem dentro dos limites de controlo e se não se verificar nenhum comportamento sistemático, concluímos que o processo estava sob controlo e que os limites de controlo obtidos se adequam a futuras monitorizações. Por outro lado, se obtivermos pontos fora dos limites de controlo, é necessário analisá-los de forma a detetar causas para tal comportamento. Se estas causas forem detetadas, os respetivos pontos são eliminados da amostra e os limites de controlo são recalculados, utilizando as restantes amostras. Este processo continua até que todos os pontos estejam dentro dos limites de controlo, sendo que a partir deste momento os limites obtidos podem ser utilizados para futuras monitorizações do processo.

---

<sup>7</sup>  $Z_{\alpha/2}$  é o valor obtido a partir da distribuição normal padronizada, tal que a probabilidade do intervalo estimado conter o parâmetro estimado é igual a  $1-\alpha$  ( $\alpha$  é o nível de significância escolhido).

### V.2.2. Cartas de Controlo EWMA

Segundo Montgomery (2009), as cartas EWMA (*Exponentially Weighted Moving Average*) são uma boa alternativa às cartas de controlo de Shewhart quando estamos interessados em detetar pequenas mudanças do processo, quando estamos perante amostras de dimensão igual a 1 ou quando a qualidade da característica a monitorizar apresenta algum afastamento em relação à distribuição Normal. Estas cartas foram propostas pela primeira vez por Roberts (1959).

A média móvel exponencialmente ponderada é dada por

$$z_i = \lambda x_i + (1 - \lambda)z_{i-1}, \quad (5.3)$$

onde  $0 < \lambda \leq 1$  é uma constante e o valor inicial (necessário para a primeira amostra em que  $i = 1$ ) é o alvo do processo, tal que:  $z_0 = \mu_0 = \bar{x}$ .

Se as observações  $x_i$  são independentes com variância  $\sigma^2$ , então a variância de  $z_i$  é

$$\sigma_{z_i}^2 = \sigma^2 \left( \frac{\lambda}{2 - \lambda} \right) \left[ 1 - (1 - \lambda)^{2i} \right]. \quad (5.4)$$

Assim, as cartas EWMA podem ser construídas representando graficamente os  $z_i$  em função da amostra  $i$ . A linha central (5.5) e os limites de controlo (5.6 e 5.7) para as cartas EWMA podem ser obtidos através das seguintes expressões:

$$CL = \mu_0 \quad (5.5)$$

$$UCL = \mu_0 + L\sigma \sqrt{\frac{\lambda}{(2 - \lambda)} \left[ 1 - (1 - \lambda)^{2i} \right]} \quad (5.6)$$

$$LCL = \mu_0 - L\sigma \sqrt{\frac{\lambda}{(2 - \lambda)} \left[ 1 - (1 - \lambda)^{2i} \right]} \quad (5.7)$$

Para este estudo, os parâmetros  $L$  e de  $\lambda$  foram escolhidos de forma a poder obter propriedades de deteção de mudanças do processo semelhantes às das cartas de controlo de Shewhart, tal que:  $L = 2,86$  e  $\lambda = 0,2$ . Repare-se que para estes parâmetros obtemos um  $ARL_0$  igual a 370,5, o que indica que em média são geradas 370,5 amostras até obtermos um ponto fora de controlo. Em Montgomery (2009), é possível ver-se o desempenho destas cartas para outros valores destes parâmetros e compará-los aos valores de desempenho das cartas Shewhart.

### V.2.3. Cartas de Controlo *Skew-Normal*

Os limites superior e inferior das cartas de controlo por variáveis de Shewhart com limites 3-sigma dependem das estimativas de  $\mu$  e  $\sigma$ . Porém, se a distribuição da variável em estudo for assimétrica, deve-se evitar recorrer aos limites 3-sigma. Neste caso, os limites inferior e superior estão à mesma distância da linha de controlo, ao contrário do que se espera de uma distribuição assimétrica. Neste sentido, é possível monitorizar os processos através de cartas de controlo recorrendo a classes de distribuições *Skew-normal* (SN). Segundo Figueiredo e Gomes (2013), mesmo em dados que possam ser considerados normais, por vezes, podem existir alguns distúrbios e a família de distribuições SN pode descrever os dados de uma forma mais fiável e robusta.

Consideremos  $X$  uma variável aleatória com distribuição SN, com parâmetro de localização  $\lambda$ , parâmetro de escala  $\delta$  e parâmetro de forma  $\alpha$ , tal que  $X \sim SN(\lambda, \delta^2, \alpha)$ . A função densidade de probabilidade é definida por

$$f_{SN}(x; \lambda, \delta, \alpha) = \frac{2}{\delta} f_N\left(\frac{x-\lambda}{\delta}\right) F_N\left(\alpha \frac{x-\lambda}{\delta}\right), \quad \{\lambda, \delta, \alpha\} \in \mathfrak{N} \quad e \quad \delta \in \mathfrak{N}^+, \quad (5.8)$$

tal que  $f_N$  é função densidade de probabilidade e  $F_N$  a função distribuição cumulativa para a distribuição Normal. (Figueiredo e Gomes, 2013)

Note-se que se  $\lambda = 0$  e  $\delta = 1$ , então estamos perante uma distribuição SN standard ( $SN(\alpha)$ ). Por outro lado, se  $\alpha = 0$  então a distribuição é Normal.

A função distribuição cumulativa é dada por

$$F_{SN}(x; \lambda, \delta, \alpha) = F_N\left(\frac{x-\lambda}{\delta}\right) - 2T\left(\frac{x-\lambda}{\delta}, \alpha\right), \quad \{\lambda, \delta, \alpha\} \in \mathfrak{N} \quad e \quad \delta \in \mathfrak{N}^+, \quad (5.9)$$

tal que  $T(h, b)$  (5.12) é a função de Owen (integral da densidade normal bivariada, com limites  $x = h$ ,  $y = 0$  e  $y = bx$ ), tabulados em Owen (1956)).

$$T(h, b) = \frac{1}{2\pi} \int_0^b \left\{ e^{-\frac{1}{2}h^2(1+x^2)} / (1+x^2) \right\} dx, \quad (b, h) \in \mathfrak{N} \times \mathfrak{N} \quad (5.10)$$

Em Azzalini (2005), é possível consultar mais desenvolvimentos sobre a distribuição SN.

Apesar da complexidade das funções anteriores, através do pacote “sn” do software estatístico *R*, é possível calcular com facilidade os valores da função densidade, da função

distribuição, os quartis, gerar números aleatórios e obter estimativas de máxima verosimilhança (EMV).

Segundo Figueiredo e Gomes (2013), para verificar se uma variável segue uma distribuição SN ou Normal, devemos recorrer ao teste de razão de verosimilhanças (5.11), sob as hipóteses  $H_0$ :  $X \sim SN(\lambda, \delta^2, \alpha = 0)$  vs.  $H_1$ :  $X \sim SN(\lambda, \delta^2, \alpha \neq 0)$ , tal que para valores elevados de  $-2 \log \Lambda$ , rejeitamos  $H_0$ , i.e., concluímos que há uma forte evidência que a distribuição  $SN(\hat{\lambda}, \hat{\delta}^2, \hat{\alpha})$  ajusta-se melhor aos dados do que a distribuição Normal.

$$\Lambda = \frac{L_N(\hat{\lambda}, \hat{\delta}, \alpha = 0)}{L_{SN}(\hat{\lambda}, \hat{\delta}, \hat{\alpha})} e^{-2 \log \Lambda} \cap \chi^2_{1gl} \quad (5.11)$$

Como não existem fórmulas explícitas das estimativas de máxima verosimilhança dos parâmetros de interesse da SN, da média e da variância, devemos recorrer a cartas de controlo por *bootstrap* para monitorizar um processo SN. Figueiredo e Gomes (2013), sugerem que se efetue os seguintes passos para a construção das cartas SN:

Fase I: Construção das cartas.

1. Considerar  $k$  amostras de dimensão  $n$  (neste caso, 1) provenientes de uma distribuição  $SN(\lambda, \delta^2, \alpha)$ .
2. Obter as EMV de  $\lambda$ ,  $\delta$  e  $\alpha$ , usando a amostra  $k \times 1$ .
3. A partir das EMV dos parâmetros de interesse, gerar, por *bootstrap*,  $B$  ( $B=100000$ ) estimativas da média.
4. Calcular os quartis  $100(p/2)$  e  $100(1-p/2)$  da distribuição das médias obtidas em (3), ou seja, os limites de controlo inferior e superior, respetivamente, tal que  $p$  é a probabilidade de uma amostra estar fora dos limites de controlo quando o processo está controlado. Assim como nos processos 3-sigma, é usual considerar  $p = 0,27\%$ .
5. Construir a Carta X-SN para a média das  $k$  amostras de dimensão 1, considerando os limites de controlo estabelecidos em (4).

Fase II: Monitorização do processo. Repare-se que, como pretendemos implementar cartas de controlo retrospetivas, esta abordagem não será efetuada.

6. Recolher amostras de dimensão 1, em intervalos regulares.

7. Se a média da amostra recolhida estiver entre os limites de controlo superior e inferior, então assume-se que o processo está sob controlo. Caso contrário, a carta assinala que o processo poderá estar fora de controlo.

#### V.2.4. Análise de capacidade do processo

A análise de capacidade do processo é uma parte vital do processo DMAIC e permite prever o quanto é que o processo suporta as tolerâncias estabelecidas, selecionar ou modificar o processo, estabelecer um intervalo adequado entre as colheitas das amostras para monitorização do processo, especificar requisitos de desempenho para novos equipamentos e reduzir a variabilidade de um processo.

Após controlar o processo, é possível avaliar a capacidade que este tem em cumprir as especificações. Estas referem-se aos limites especificados para o processo, por exemplo, no caso do T proc o limite superior de especificação é igual a 120 dias. Para um processo cuja distribuição da característica da qualidade seja Normal, podemos estimar a capacidade do processo através da probabilidade de obter um ponto acima do limite superior de especificação ou abaixo do limite inferior de especificação, tal que

$$P(x > USL) = 1 - \Phi\left(\frac{USL - \mu}{\sigma}\right) \text{ ou } P(x < LSL) = \Phi\left(\frac{\mu - LSL}{\sigma}\right), \quad (5.12)$$

respectivamente, e  $\Phi$  é a função distribuição cumulativa para uma distribuição Normal (0,1).

Adicionalmente, também é possível avaliar a capacidade que um processo tem em cumprir as especificações superiores e inferiores, através do rácio da capacidade do processo (PCR – *Process Capability Ratio*):

$$C_p = \frac{USL - LSL}{6\sigma}, \quad (5.13)$$

onde  $USL$  é o limite superior de especificação e  $LSL$  o limite inferior de especificação.

Porém, o índice  $C_p$  não tem em consideração a média estimada e, se o processo não estiver centrado, esta medida pode não ser a mais adequada. Neste sentido, quando pretendemos verificar a capacidade do processo em relação à média  $\mu$  estimada, recorremos a

$$C_{pk} = \min\{C_{pkl}, C_{pku}\} = \min\left\{\frac{\mu - LSL}{3\sigma}, \frac{USL - \mu}{3\sigma}\right\} \quad (5.14)$$

Repare-se que no caso deste estudo, apenas recorremos ao  $C_{pku}$  para medir a capacidade dos diferentes processos relativamente à média, pois apenas consideramos os limites de especificação superiores.

Se  $C_{pku} = 1$ , então a média do processo está dentro dos respetivos limites de especificação e concluímos que 1350 amostras num milhão não cumprem as especificações superiores. Repare-se que o número de amostras (por milhão) que saem fora do limite de especificação é calculado a partir das expressões (5.12) e (5.14), tal que

$$P(x > USL) \times 10^6 = \left[1 - \Phi\left(\frac{USL - \mu}{\sigma}\right)\right] \times 10^6 = \left[1 - \Phi(3 \times C_{pku})\right] \times 10^6.$$

Segundo Montgomery (2009), o valor mínimo para este índice deve ser igual a 1,25, que indica que apenas 88 amostras num milhão estão fora do limite superior de especificação. Por outro lado, se  $C_{pku} = 0$  a média do processo é igual ao limite superior de especificação, se  $C_{pku} < 0$  a média do processo está acima do limite superior de especificação e se  $C_{pku} < -1$  todo o processo está fora do limite de especificação.

Para um processo cuja distribuição da característica da qualidade se ajuste a uma SN, é necessário ajustar as expressões (5.12) e (5.14) à distribuição SN  $(\lambda, \delta^2, \alpha)$ , tal que

$$P(x > USL) = 1 - \Phi_{SN(\lambda, \delta^2, \alpha)}(USL) \quad \text{e} \quad P(x < LSL) = \Phi_{SN(\lambda, \delta^2, \alpha)}(LSL) \quad (5.15)$$

e

$$C_{pku} = \frac{USL - \mu}{a - b} \quad \text{e} \quad C_{pku} = \frac{\mu - LSL}{a - b}, \quad (5.16)$$

onde  $a$  é igual ao quartil  $100(1-p/2)$  da distribuição das médias obtidas a partir das B estimativas da média e  $b$  o quartil  $100(p/2)$ .

### V.3. Caracterização da amostra

Para o estudo dos diferentes tempos, foi analisada uma amostra constituída por 1095 alunos diplomados que ingressaram na Universidade de Évora entre o ano letivo de 2007/2008 e o ano letivo de 2012/2013, tendo sido seguidos até Julho de 2013.

Numa primeira análise ao T proc, concluímos que metade destes tempos excede os 122 dias. O valor mínimo obtido é igual a 4 dias e o máximo igual a 808 dias. Para além disso, a média é igual a 141 dias e o desvio padrão é igual a 91,9 dias (Tabela 5.1). Assim, uma análise exploratória a esta variável permitiu identificar tempos de procedimentos excessivamente elevados que estão muito longe de cumprir o regulamento interno desta Universidade. Segundo este regulamento, o tempo de procedimentos não deverá exceder os 120 dias.

Tendo em conta o objetivo de monitorizar este processo, optámos por implementar cartas de controlo estatístico apenas aos tempos de procedimentos inferiores ou iguais a 180 dias, resultando numa amostra de 853 alunos diplomados. Para os alunos desta amostra, concluímos que 55% pertencem à Escola de CS, 44% à Escola de CT e os restantes à Escola de A. Também verificamos que 3% das observações são relativas ao ano de conclusão de 2009, 17% de 2010, 25% de 2011, 34% de 2012 e restantes de 2013.

Já os tempos excessivamente elevados representam 22% da amostra inicial, ou seja, 242 alunos cujos T proc excederam os 180 dias. Para os alunos desta amostra, concluímos que 65% pertencem à Escola de CS, 31% à Escola de CT e os restantes à Escola de A. Obtivemos 15% de registo em 2010, 34% em 2011, 37% em 2012 e 14% em 2013, sendo que em 2009 não se verificou qualquer registo.

Uma análise ao histograma, à caixa de bigodes e ao qqPlot da Figura 5.1 e ao valor de prova do teste de normalidade de Kolmogorov-Smirnov<sup>8</sup> (KS) (Tabela 5.1), permitiram concluir que o T proc não tem distribuição normal e que a curva de distribuição desta variável é acentuadamente assimétrica. O mesmo se concluiu para os T proc excessivamente elevados (Figura 5.2).

Já para os T proc inferiores ou iguais a 180 dias, concluímos que sua a distribuição é simétrica e que se aproxima um pouco mais da distribuição normal (Figura 5.3). No caso destes tempos, recorremos ao método de Box-Cox para proceder a uma transformação da variável e tentámos encontrar uma distribuição que se ajustasse aos dados, nomeadamente uma distribuição

<sup>8</sup> O teste de KS tem por base a análise da proximidade da função distribuição populacional  $F_0(x)$  admitida em  $H_0$  e a função distribuição empírica ou da amostra  $F_n^*(x)$ . A estatística de teste é dada por  $D_n = \sup_x |F_0(x) - F_n^*(x)|$  e compara-se o valor da estatística de teste com o valor crítico tabulado para o nível de significância de 5%. Se  $D_n > D_{tab}$ , então rejeita-se a hipótese nula de que a variável segue uma distribuição normal.

Weibull, Gamma, Exponencial ou Log-normal. Contudo, para a transformação sugerida obtivemos um valor de prova do teste de KS inferior a 5% (Tabela 5.1) e concluímos que o T proc não segue uma distribuição de Weibull, Gamma, Exponencial ou log-normal.

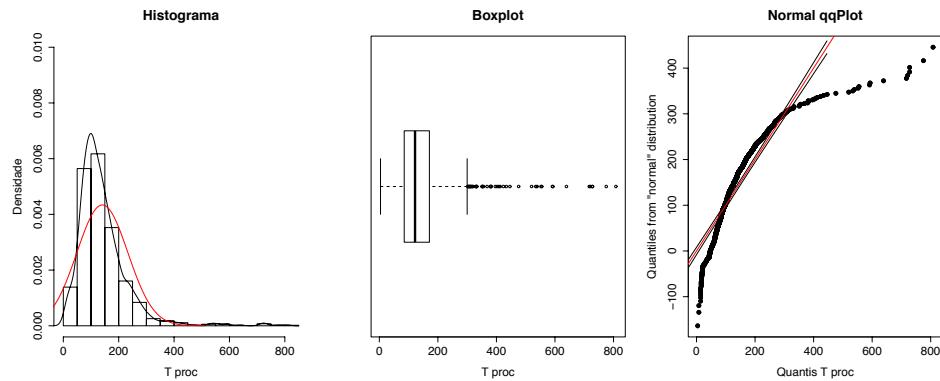


Figura 5.1: Histograma, Caixa de Bigodes e qqPlot para a variável T proc.

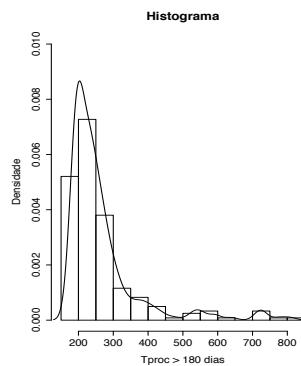


Figura 5.2: Histograma para a variável T proc superior a 180 dias.

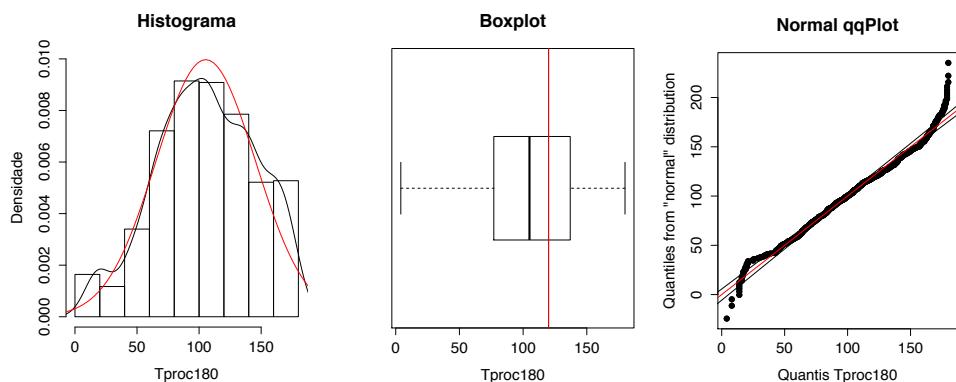


Figura 5.3: Histograma, Caixa de Bigodes e qqPlot para a variável T proc inferior ou igual a 180 dias.

**Tabela 5.1: Estatística descritiva e testes de normalidade para o T proc, T proc ( $\leq 180$  dias) e T proc ( $\leq 180$  dias) transformado.**

	T proc	T proc ( $\leq 180$ dias)	T proc <sub>(<math>\leq 180</math> dias)</sub> Transf = (T proc <sub>(<math>\leq 180</math> dias)</sub> ) <sup>1,2</sup>
Mínimo	4	4	5,3
1º Quartil	85	77	183,6
Mediana	122	105	266,3
3º Quartil	171	137	366,5
Máximo	808	180	508,5
Média	141	105,4	272,5
Desvio Padrão	91,9	40,0	119,7
Coeficiente de variação	65%	38%	44%
Coeficiente de assimetria	2,7	-0,2	<0,001
Valor p (teste D'Agostino <sup>9</sup> )	<0,001	0,15	0,99
Coeficiente de achatamento	12,6	-0,6	-0,7
Valor p (teste Anscombe <sup>10</sup> )	<0,001	<0,001	<0,001
Valor p (teste KS)	<0,001	0,001	0,001

Tendo em conta que os T sac, T formação do júri, T reunião e T provas são tempos parciais do T proc (Figura 1.1, pág. 4), optámos por também estudar estes tempos considerando apenas os 853 alunos diplomados. Repare-se que, segundo o regulamento interno desta Universidade, o T sac não deverá exceder 5 dias, os T formação do júri e T reunião não deverão exceder 30 dias e o T provas não deverá ultrapassar 60 dias.

Uma análise à distribuição destes tempos permitiu concluir que nenhum tem um comportamento que se aproxima de uma distribuição normal e que as curvas de distribuição destas variáveis são acentuadamente assimétricas (Figuras 5.4 a 5.7). Também podemos concluir que alguns valores do T sac ( $\leq 180$  dias) são muito elevados relativamente aos prazos estabelecidos pelo regulamento interno desta Universidade, pois metade desta amostra excede os 10 dias. Quanto aos restantes tempos, concluímos que para metade da amostra o T formação do júri ( $\leq 180$  dias) é superior a 33 dias e o T reunião ( $\leq 180$  dias), assim como o T provas ( $\leq 180$  dias), são superiores a 21 dias (Tabela 5.2).

<sup>9</sup> O teste de D'Agostino avalia o afastamento à normal em termos de simetria.

<sup>10</sup> O teste Anscombe-Glynn avalia o afastamento à normal em termos de curtose.

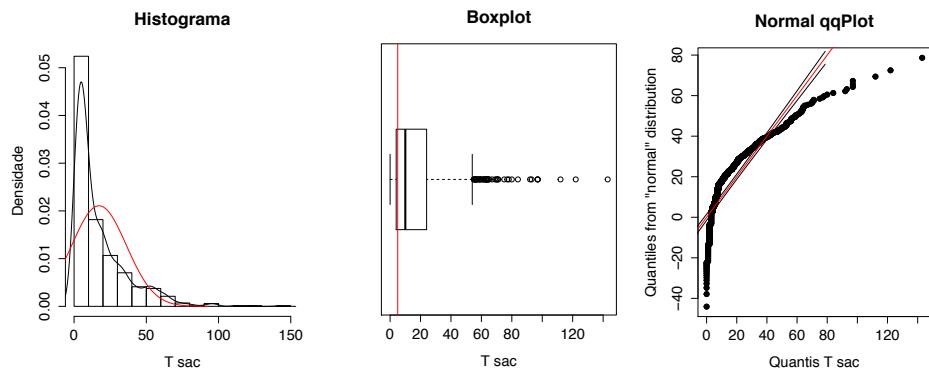


Figura 5.4: Histograma, Caixa de Bigodes e qqPlot para a variável T sac ( $\leq 180$  dias).

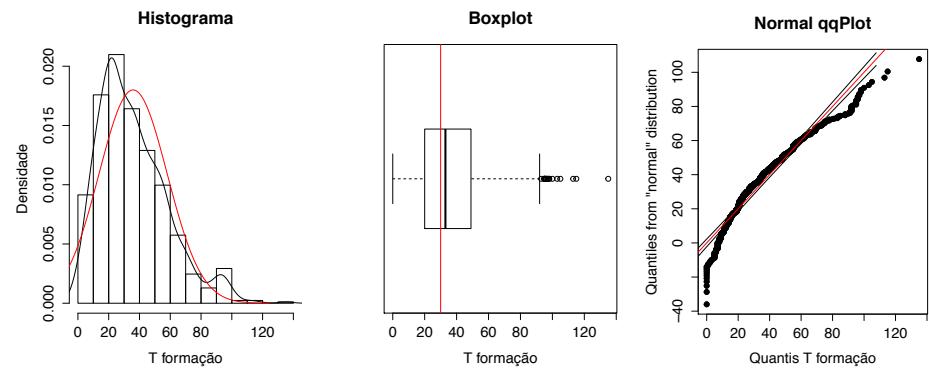


Figura 5.5: Histograma, Caixa de Bigodes e qqPlot para a variável T formação do júri ( $\leq 180$  dias).

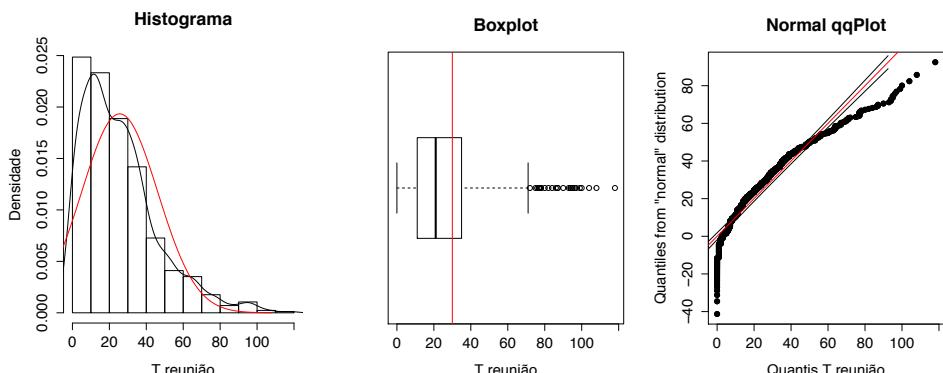


Figura 5.6: Histograma, Caixa de Bigodes e qqPlot para a variável T reunião ( $\leq 180$  dias).

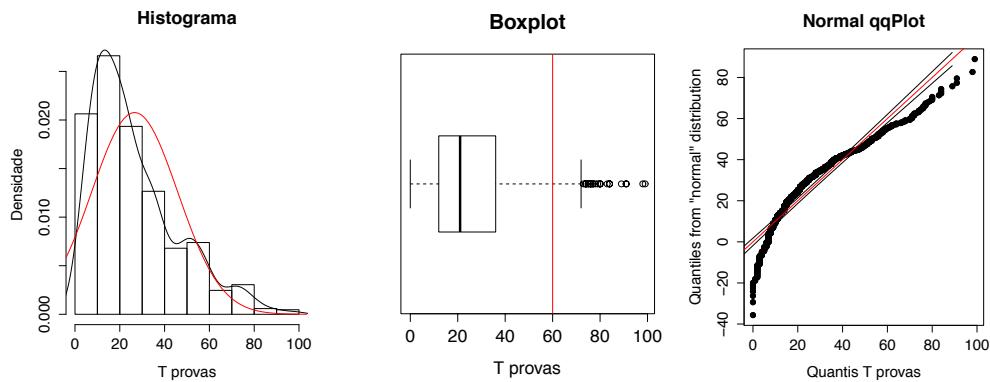


Figura 5.7: Histograma, Caixa de Bigodes e qqPlot para a variável T provas ( $\leq 180$  dias).

Tabela 5.2: Estatística descritiva e testes de normalidade para o T sac, Tsac ( $\leq 180$  dias), T formação do júri, T formação do júri ( $\leq 180$  dias), T reunião, T reunião ( $\leq 180$  dias), T provas e T provas ( $\leq 180$  dias).

	T sac	T sac ( $\leq 180$ )	T formação do júri	T formação do júri ( $\leq 180$ )	T reunião	T reunião ( $\leq 180$ )	T provas	T provas ( $\leq 180$ )
Mínimo	0	0	0	0	0	0	0	0
1º Quartil	5	4	21	20	12	11	14	12
Mediana	12	10	36	33	26	21	24	21
3º Quartil	29	24	55,5	49	45	35	46	36
Máximo	658	143	265	135	610	118	665	99
Média	28,7	17,3	43,3	35,9	34,5	25,6	34,4	26,6
Desvio Padrão	61,5	18,9	32,4	22,2	37,6	20,6	36	19,2
Coeficiente de variação	214%	109%	75%	62%	109%	80%	105%	72%
Coeficiente de assimetria	6,3	2,0	2	0,9	5,1	1,2	6,3	1,1
Valor p (teste D'Agostino)	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001
Coef. de achatamento	48,4	5,5	7,2	0,8	56,3	1,7	89,1	0,7
Valor p (teste Anscombe)	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001	0,001
Valor p (teste KS)	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001

Para qualquer um dos anos de conclusão considerados (Figuras 5.8 a 5.12), o tempo de formação do júri e o tempo dos SAC são aqueles que mais contribuem para os T proc ( $\leq 180$  dias) mais elevados. Por exemplo, no ano de 2009, a partir de Julho parece haver um aumento dos T proc, que se deve em grande parte ao T formação do júri e ao T sac (Figura 5.8). É de referir que no mês de Junho obtivemos valores baixos do T proc, sendo que se registaram de T reunião e T provas inferiores a 1 dia. Estes tempos parecem ser baixos demais tendo em conta a característica dos mesmos. Ou seja, para alguns casos, concluímos que decorreu apenas um dia entre a data de homologação e a data do despacho liminar e entre a data do despacho e a data das provas.

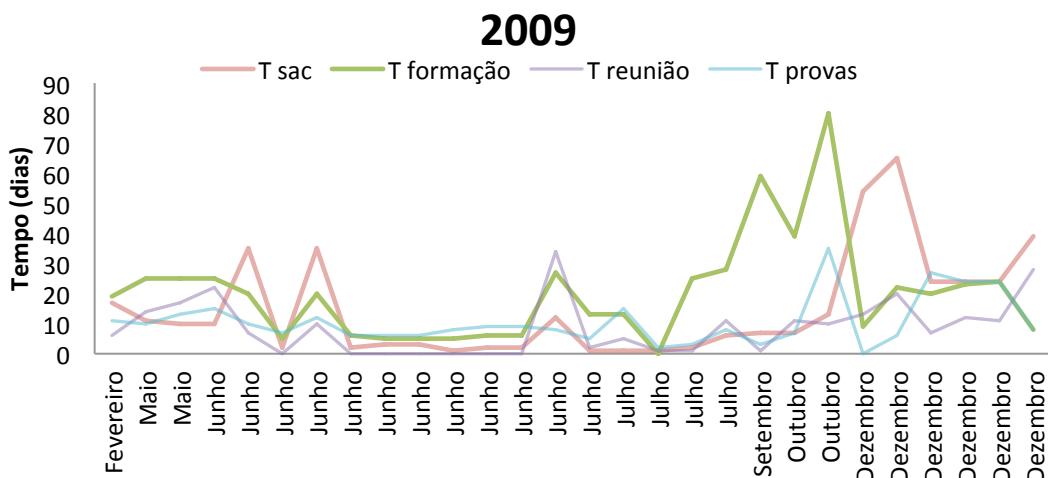


Figura 5.8: Distribuição dos tempos parciais (T sac, T formação do júri, T reunião e T provas) que compõem o T proc ( $\leq 180$  dias), em função dos meses do ano de 2009.

Quanto ao ano de 2010, concluímos que os T proc tendem a aumentar até Março (apesar de haver uma grande instabilidade no mês de Janeiro), que decrescem após esse mês, mas que voltam a aumentar entre Julho e Setembro. Também se pode verificar que o processo parece ser muito instável, pois, por exemplo, no mês de Novembro tanto podemos ter valores de T proc muito baixos como tempos acima dos 120 dias (Figura 5.9). Por outro lado, também neste ano concluímos que os maiores valores de T proc devem-se em grande parte ao T formação do júri.

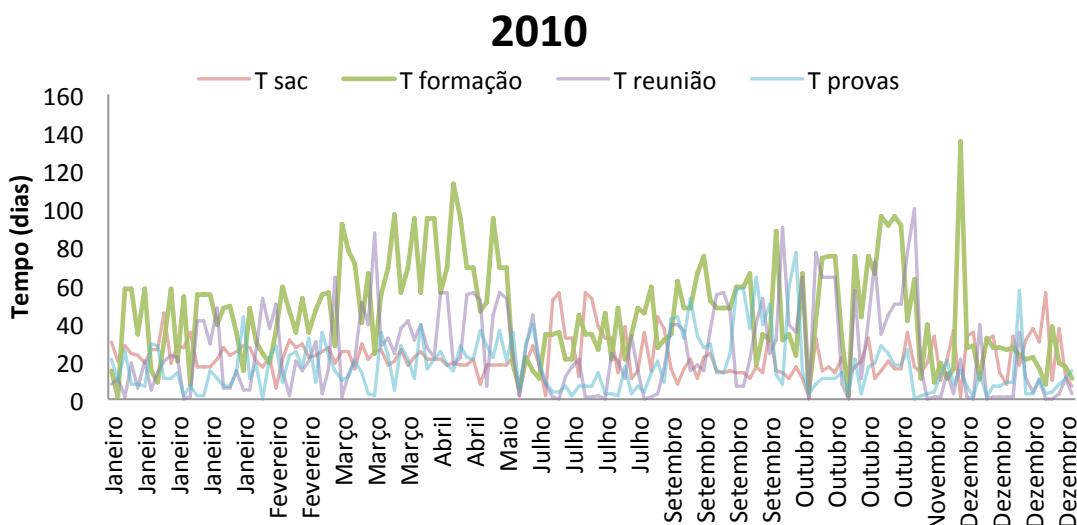
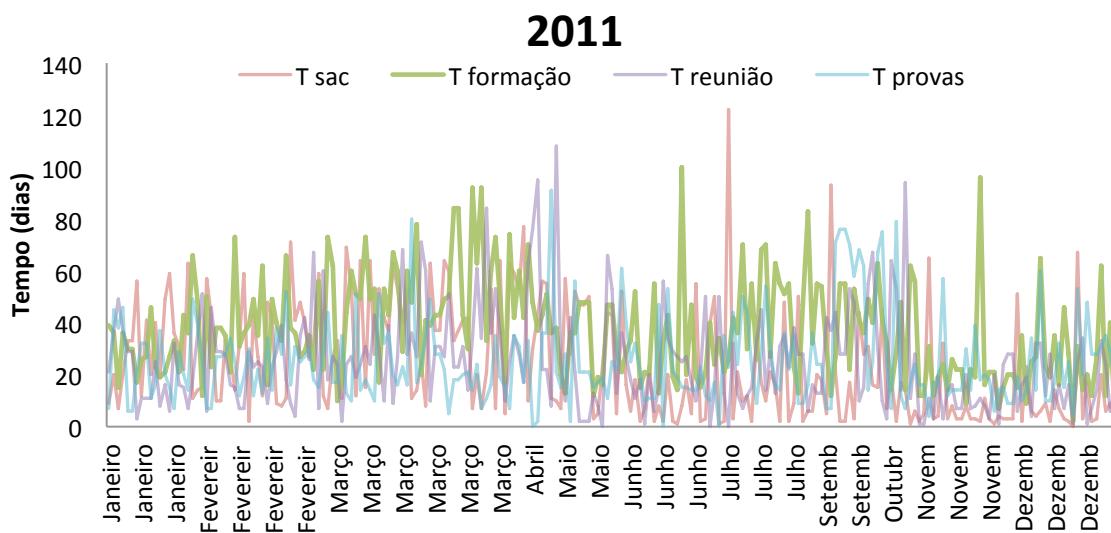


Figura 5.9: Distribuição dos tempos parciais (T sac, T formação do júri, T reunião e T provas) que compõem o T proc ( $\leq 180$  dias), em função dos meses do ano de 2010.

Nos primeiros meses de 2011 registaram-se T proc superiores aos do ano anterior (Figura 5.10). Porém a tendência parece ser a mesma, pois até ao mês de Março verifica-se um aumento deste tempo, seguindo-se de uma grande instabilidade dos tempos nos meses seguintes. Neste caso, os tempos elevados devem-se, não só ao T formação do júri, mas também à grande instabilidade do T sac e do T provas.

Apesar de se verificar uma certa instabilidade dos T sac no ano de 2012 (Figura 5.11), para a maioria dos casos, é possível concluir que os tempos responsáveis pelos valores elevados do T proc são os T formação do júri, T reunião e T provas. Até Março de 2013 (Figura 5.12), os T sac são os principais responsáveis pelos valores elevados obtidos para o T proc. Nos meses seguintes, concluímos que os T sac reduzem (apesar de se manter alguma instabilidade), sendo os restantes tempos parciais os responsáveis pelos T proc mais elevados.

Note-se que em qualquer um destes anos se verifica a mesma instabilidade do processo e de alguma forma parece haver uma questão sazonal, pois até Março os T proc tendem a aumentar e voltamos a ter tempos mais elevados no final dos anos letivos.



**Figura 5.10: Distribuição dos tempos parciais (T sac, T formação do júri, T reunião e T provas) que compõem o T proc ( $\leq 180$  dias), em função dos meses do ano de 2011.**

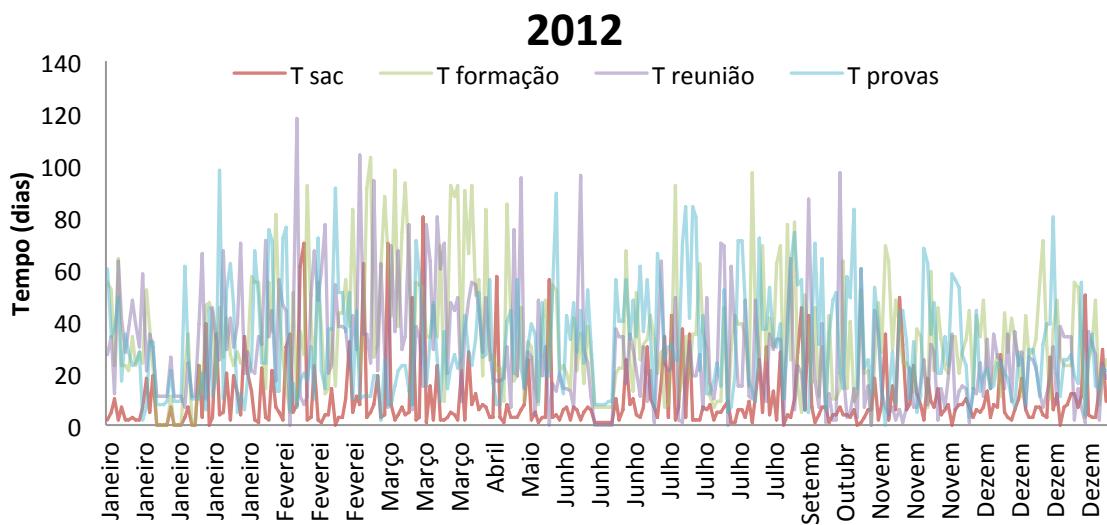


Figura 5.11: Distribuição dos tempos parciais (T sac, T formação do júri, T reunião e T provas) que compõem o T proc ( $\leq 180$  dias), em função dos meses do ano de 2012.

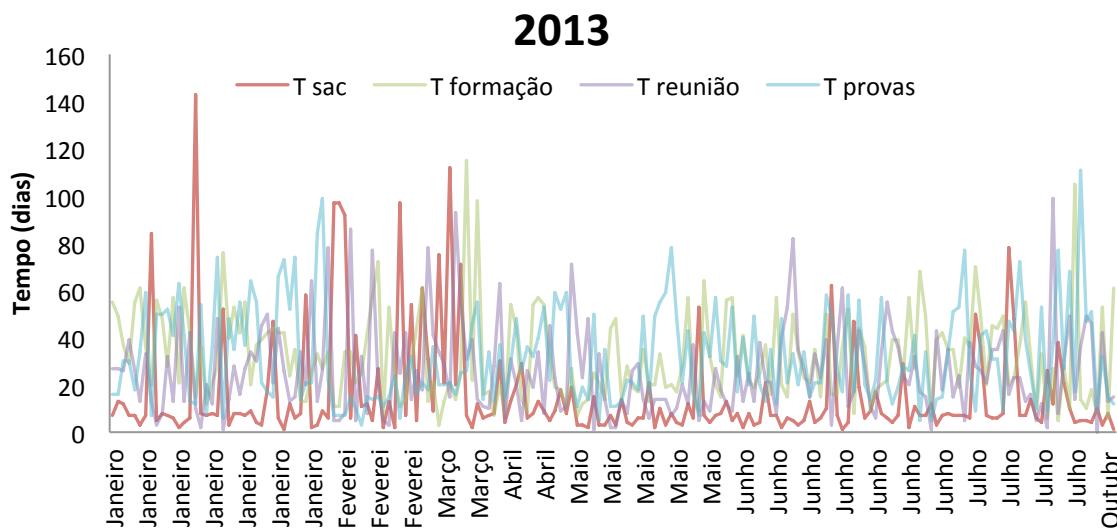


Figura 5.12: Distribuição dos tempos parciais (T sac, T formação do júri, T reunião e T provas) que compõem o T proc ( $\leq 180$  dias), em função dos meses do ano de 2013.

Finalmente, fazemos uma breve caracterização aos tempos de espera pela emissão de um diploma de grau e de curso. Para a análise do TD grau consideramos uma amostra com 3732 observações e para o TD curso uma amostra de dimensão igual a 692. Uma primeira análise a estes tempos (Tabela 5.3) permitiu concluir que 75% da amostra do TD grau e do TD curso não excedeu os 11 e 8 dias, respetivamente. Tendo em conta as características destes processos e considerando

que o regulamento interno desta Universidade prevê que estes tempos não excedam 30 dias, apenas consideramos as amostras relativas aos tempos inferiores ou iguais a 40 dias para monitorizar os processos em causa.

Note-se que os tempos superiores a 40 dias representam 3% da amostra inicial do TD grau e 5% da amostra do TD curso. No primeiro caso, a média destes tempos é igual a 87,4 dias e o desvio padrão é igual a 62,7 dias, sendo que para metade desta amostra o TD grau ( $> 40$  dias) excede os 65 dias. A média do TD curso ( $> 40$  dias) é igual a 109,9 dias, o desvio padrão igual a 73,3 dias e metade destes tempos são superiores a 82 dias.

No caso do TD grau inferior ou igual a 40 dias, obtivemos uma amostra com 3622 registos, para os quais 9% são relativos ao ano de 2010, 35% de 2011, 35% de 2012 e os restantes de 2013. Já para o TD curso inferior ou igual a 40 dias, consideramos uma amostra com 656 registos, sendo que 5% são relativos ao ano de 2010, 48% de 2011, 26% de 2012 e os restantes de 2013.

Concluímos que para metade dos registos do TD grau ( $\leq 40$  dias) este tempo de espera não excede 6 dias, obtivemos uma média igual a 7,3 dias e um desvio padrão igual a 6,6 dias. Para o TD curso ( $\leq 40$  dias), concluímos que a média do tempo de espera pela emissão de um diploma é muito semelhante (6,3 dias) à média do TD grau, porém obtivemos uma mediana inferior (4 dias) (Tabela 5.3).

**Tabela 5.3: Estatística descritiva e testes de normalidade para o TD grau, TD grau ( $\leq 40$  dias), TD curso e TD curso ( $\leq 40$  dias).**

	TD grau	TD grau ( $\leq 40$ dias)	TD curso	TD curso ( $\leq 40$ dias)
Mínimo	0	0	0	0
1º Quartil	3	3	3	3
Mediana	6	6	5	4
3º Quartil	11	10	8	7
Máximo	369	40	290	40
Média	9,7	7,3	11,7	6,3
Desvio Padrão	18,4	6,6	29	6,6
Coeficiente de variação	190%	90%	248%	105%
Coeficiente de assimetria	9,6	1,7	6,2	2,4
Valor p (teste D'Agostino)	<0,001	<0,001	<0,001	<0,001
Coeficiente de achatamento	134,5	3,7	45,2	6,5
Valor p (teste Anscombe)	<0,001	<0,001	<0,001	<0,001
Valor p (teste KS)	<0,001	<0,001	<0,001	<0,001

## V.4. Análise dos dados

### V.4.1. Tempo entre a entrega e discussão da dissertação

Após uma análise à distribuição dos T proc inferiores ou iguais a 180 dias, optámos por construir cartas de controlo EWMA, pois estas cartas são mais robustas para desvios à normalidade (Tabela 5.1) e são mais adequadas a amostras de dimensão 1. Numa primeira análise à carta de controlo da Figura 5.13, podemos concluir que este processo não está sob controlo, pois obtivemos valores da estatística EWMA superiores e inferiores aos UCL e LCL, respetivamente.

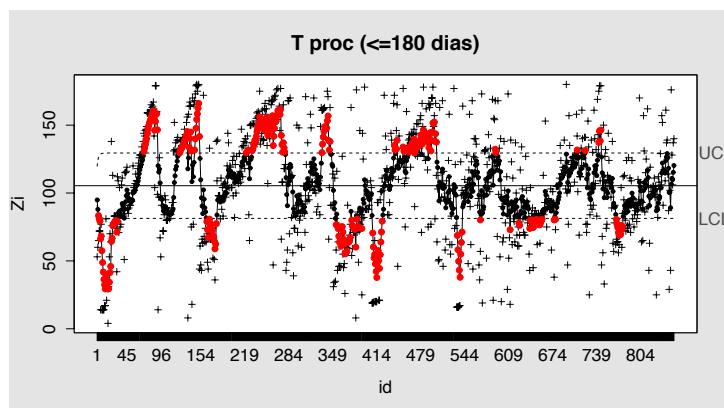


Figura 5.13: Carta de controlo EWMA para o T proc ( $\leq 180$  dias).  $CL = 105,4$  dias,  $sd = 25,3$  e  $k = 853$ .

Para compreender melhor o processo e para tentar detetar causas responsáveis pela variação do mesmo, estratificamos os tempos inferiores a 180 dias pelos diferentes anos de conclusão, escolas, departamentos e por alguns cursos de Mestrado. Para efetuar esta análise tivemos sempre em consideração a distribuição dos tempos em causa, para que a escolha das cartas a implementar fosse a mais adequada. Assim, para a escolha das Cartas EWMA ou Cartas X-SN tivemos em consideração os resultados dos testes de KS e SN e os parâmetros estimados das distribuições SN (Tabela 1, Anexo 2).

Ao estratificar pelos anos de conclusão (Figura 5.14), verificamos que há uma sazonalidade inerente a este processo. Para além disso, para todos os anos de conclusão, obtivemos valores da estatística EWMA superiores ao UCL e inferiores ao LCL, por isso em nenhum dos casos podemos considerar que o processo está sob controlo. Concluímos também que a média estimada é inferior no ano de 2009 ( $LC = 54,1$  dias) e que nos anos de 2010 e 2011 obtivemos mais pontos fora dos limites de controlo e médias estimadas mais elevadas. Em 2010, 2011, 2012 e 2013 as médias

estimadas são iguais a 110,5, 112,1, 102,2 e 106,2 dias, respetivamente. Através do teste de Kruskal Walis, apenas se verificam diferenças entre as médias estimadas no ano de 2009 e nos restantes anos (valor  $p < 0,001$ ).

Ao estratificar o referido tempo pelas diferentes escolas (Figura 5.15), concluímos que apenas para a Escola de A é que o processo está sob controlo. Contudo, a sua capacidade em cumprir as especificações superiores é muito baixa ( $C_{pku} = 0,05$ , i.e., 443832 amostras num milhão estão acima do limite superior de especificação) e a probabilidade de termos um ponto acima dos 120 dias é elevada (Probabilidade = 0,44). Já quando se consideram as escolas de CS e CT, concluímos que estes processos estão fora de controlo. Para tentar compreender melhor o processo inerente a estas duas escolas, estratificamos as amostras pelos diferentes departamentos.

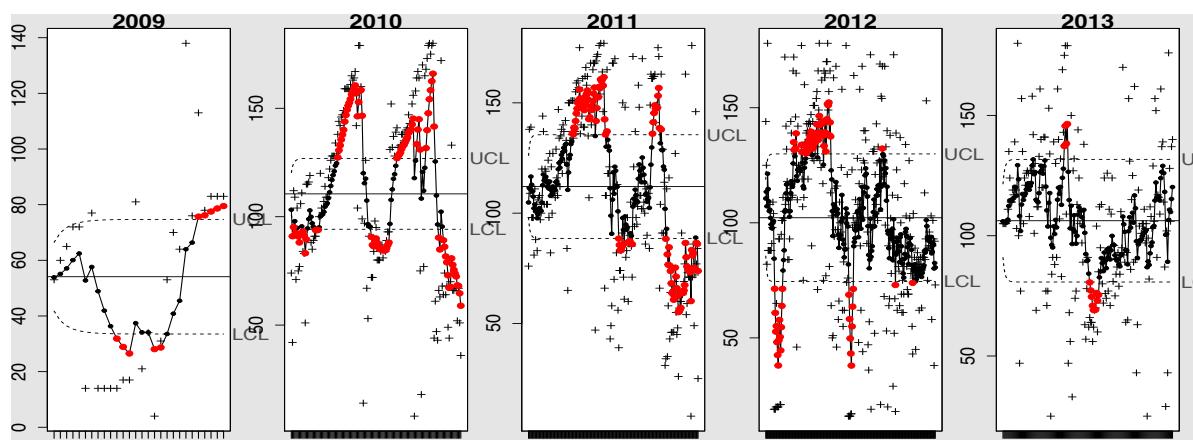


Figura 5.14: Cartas de controlo EWMA para os anos de conclusão: 2009 ( $CL = 54,1$  dias,  $sd = 21,6$  e  $k = 28$ ), 2010 ( $CL = 110,5$  dias,  $sd = 17,1$  e  $k = 147$ ), 2011 ( $CL = 112,1$  dias,  $sd = 24,6$  e  $k = 217$ ), 2012 ( $CL = 102,2$  dias,  $sd = 29,1$  e  $k = 286$ ) e 2013 ( $CL = 106,2$  dias,  $sd = 26,7$  e  $k = 175$ ).

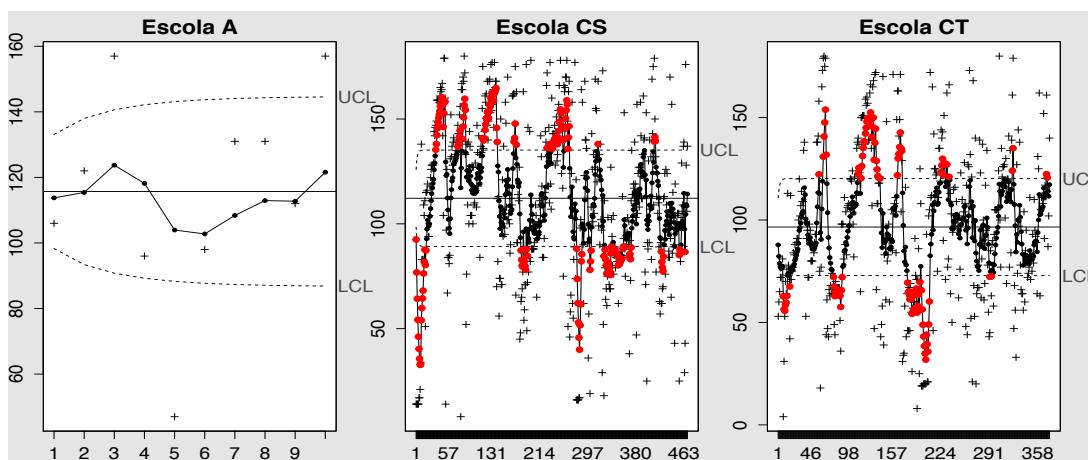


Figura 5.15: Cartas de controlo EWMA para a Escola de A ( $CL = 115,7$  dias,  $sd = 30,4$  e  $k = 10$ ), de CS ( $CL = 112,2$  dias,  $sd = 24,1$  e  $k = 469$ ) e CT ( $CL = 96,6$  dias,  $sd = 24,8$  e  $k = 374$ ).

Uma análise às cartas de controlo das Figuras 5.16 e 5.17, permitiu concluir que os processos estão controlados quando consideramos os departamentos de Economia, Matemática, Paisagem, Ambiente e Ordenamento, Química e Zootecnia. Apesar destes processos estarem sob controlo, obtivemos índices da capacidade do processo muito baixos. Por exemplo, para o departamento de Economia obtivemos um  $C_{pku}$  igual a 0,11 (372846 amostras num milhão estão acima do limite superior de especificação) e uma probabilidade de um ponto exceder os 120 dias elevada ( $P = 0,37$ ). Para os restantes casos, obtivemos um  $C_{pku} = 0,16$  e Probabilidade = 0,32 para o departamento de Matemática,  $C_{pku} = 0,32$  e Probabilidade = 0,17 para o departamento Paisagem, Ambiente e Ordenamento,  $C_{pku} = 0,05$  e Probabilidade = 0,44 para o departamento Química e  $C_{pku} = 0,20$  e Probabilidade = 0,27 para o departamento Zootecnia.

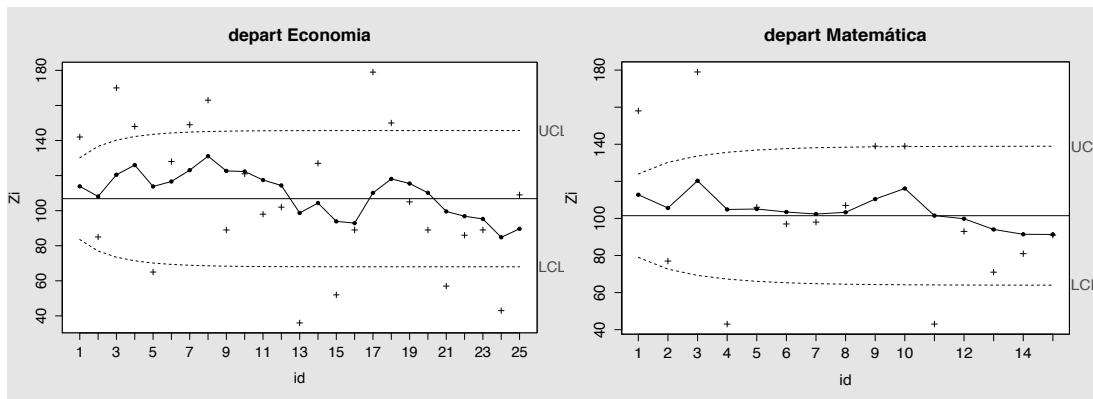


Figura 5.16: Cartas de controlo EWMA para o departamento de Economia ( $CL = 106,8$  dias,  $sd = 40,7$  e  $k = 25$ ) e de Matemática ( $CL = 101,5$  dias,  $sd = 39,3$  e  $k = 15$ ).

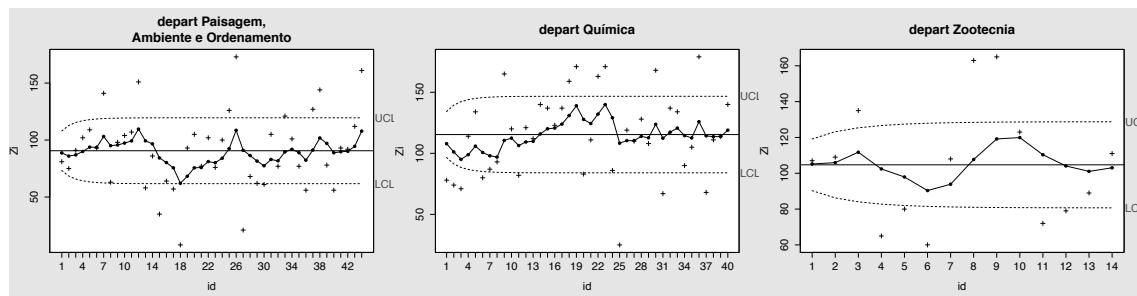


Figura 5.17: Cartas de controlo EWMA para o departamento de Paisagem, Ambiente e Ordenamento ( $CL = 90,6$  dias,  $sd = 30,3$  e  $k = 44$ ), de Química ( $CL = 115,4$  dias,  $sd = 32,9$  e  $k = 40$ ) e de Zootecnia ( $CL = 104,7$  dias,  $sd = 25,2$  e  $k = 14$ ).

Adicionalmente, verificamos que também para os departamentos de Física e Psicologia o processo está controlado. Nestes casos, recorremos a cartas de controlo X-SN, pois as distribuições dos tempos em causa ajustam-se a uma distribuição SN (Tabela 1, Anexo 2).

Repare-se na importância da distribuição em causa, pois se considerássemos Cartas EWMA ou Cartas X de Shewhart para o departamento de Física concluiríamos que o processo não estava sob controlo (Figura 5.18). Em qualquer uma das cartas, concluímos que existem algumas amostras cujos  $T_{proc}$  são muito mais baixos do que seria de esperar, inferiores a 22 dias, o que evidencia uma redução da média estimada. Repare-se que o registo destes tempos ocorreu no mês de Janeiro de 2012 e para o curso de Ciências da Terra, da Atmosfera e do Espaço (a partir da amostra 13). Apesar de apenas a carta EWMA apresentar pontos abaixo do limite inferior de controlo, a carta X indica que temos 9 pontos consecutivos abaixo da linha central, situação esta muito rara de ocorrer (Probabilidade = 0,002) e que também traduz uma diminuição da média estimada.

Ao considerar a Carta X-SN da Figura 5.18, concluímos que apesar deste processo apresentar uma baixa probabilidade de ter uma amostra acima dos 120 dias, obtivemos um valor um  $C_{pku}$  muito baixo ( $C_{pku(SN)} = 0,33$  e Probabilidade<sub>SN</sub> = 0,07).

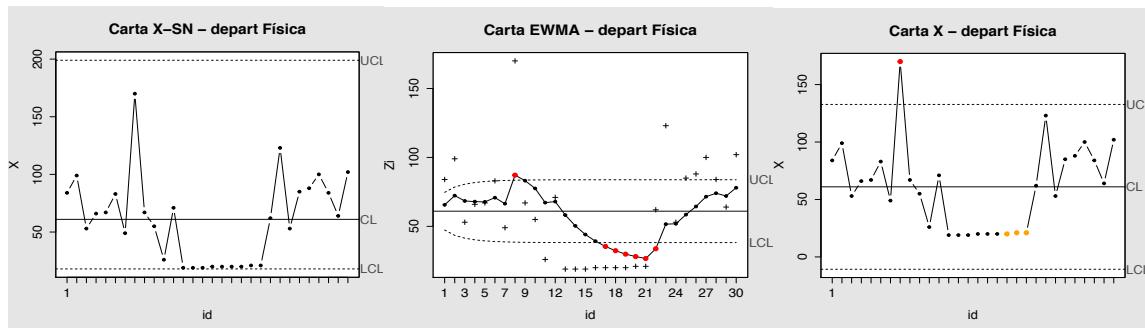


Figura 5.18: Cartas de controlo para o departamento de Física ( $CL = 61$  dias,  $sd = 23,9$  e  $k = 30$ ): Carta X-SN ( $LCL = 18$  e  $UCL = 199$ ); Carta EWMA; Carta X ( $LCL = -10,7$  e  $UCL = 132,7$ ).

Relativamente ao departamento de Psicologia, apenas pudemos considerar as amostras relativas ao curso de Psicologia. Repare-se que apesar do processo estar sob controlo, uma análise à Carta X-SN deste curso (Figura 5.19) permitiu identificar dois períodos distintos, até ao ano letivo de 2011/2012 e depois de 2011/2012. Efetivamente, quando se considera o primeiro período, a média estimada ( $CL = 144,3$ ) é superior à do segundo período ( $CL = 107,5$ ) (valor  $p$  do teste de Wilcoxon  $< 0,001$ ). Ao investigar as causas responsáveis por tal comportamento, no final do ano letivo de 2011/2012, verificaram-se alterações relacionadas com os processos de formação do júri que conduziram a uma redução do  $T_{proc}$ . Como seria de esperar, para o segundo período deste curso verifica-se uma maior capacidade do processo cumprir as especificações superiores ( $C_{pku} = 0,16$  e Probabilidade = 0,31) do que para o primeiro período ( $C_{pku(SN)} = -0,17$  e Probabilidade<sub>SN</sub> =

0,82). Porém, em ambos os casos, a capacidade do processo cumprir as especificações é muito baixa.

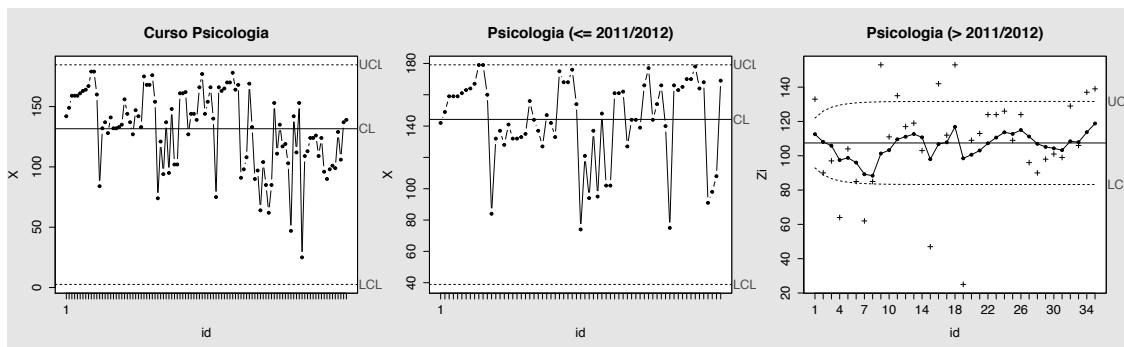


Figura 5.19: Carta X-SN para o curso de Psicologia ( $CL = 131,6$  dias,  $sd = 20,3$ ,  $LCL = 2,6$ ,  $UCL = 184,6$  e  $k = 102$ ), Carta X-SN para o primeiro período (até ao ano letivo de 2011/2012) do curso de Psicologia ( $CL = 144,3$  dias,  $sd = 17,5$ ,  $LCL = 38,9$ ,  $UCL = 179,1$  e  $k = 67$ ) e carta EWMA para o segundo período (depois de 2011/2012) do curso de Psicologia ( $CL = 107,5$  dias,  $sd = 25,4$  e  $k = 35$ ).

Quanto aos restantes departamentos, concluímos que estes processos estão fora de controlo. Nesta situação temos o departamento de Gestão, História, Pedagogia da Educação, Sociologia, Biologia, Desporto e Saúde e Informática. Tendo em conta a grande variabilidade destes processos, optamos por estratificar os referidos departamentos por alguns cursos e, eventualmente, por períodos distintos.

Ao estratificar o departamento de Gestão pelos diferentes cursos, concluímos que apenas para o curso de Gestão é que o processo não está sob controlo. Apesar das médias estimadas dos diferentes cursos serem idênticas (aproximadamente 118 dias), é no curso de Gestão que se verificam alguns pontos fora de controlo (Figura 5.20). A carta do curso de gestão indica que as amostras 31 e 32 estão acima do limite superior de controlo, evidenciando um aumento dos tempos médios nos meses de Março e Abril de 2011. Por outro lado, a partir da amostra 60 (relativa ao mês de Dezembro de 2012) houve uma diminuição da média do T proc, o que se reflete numa amostra abaixo do limite inferior de controlo.

Repare-se que apesar de não termos pontos fora dos limites de controlo quando consideramos a carta dos restantes cursos deste departamento, este processo apresenta uma baixa capacidade em cumprir as especificações superiores ( $C_{pk} = 0,02$  e Probabilidade = 0,47).

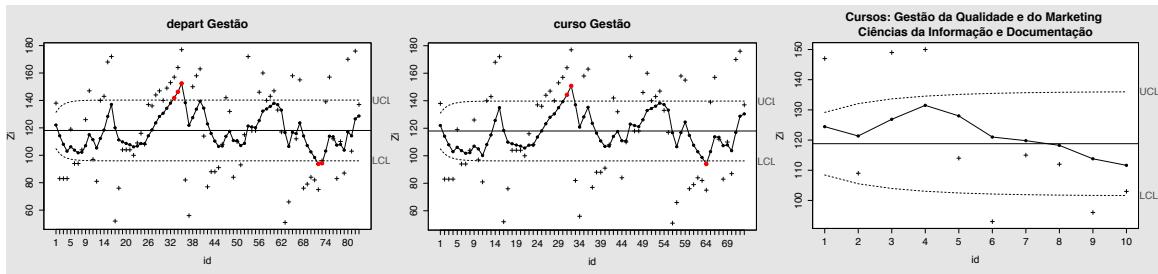


Figura 5.20: Cartas de controlo EWMA para o departamento de Gestão ( $CL = 118,1$  dias,  $sd = 23,2$  e  $k = 83$ ), o curso de Gestão ( $CL = 118$  dias,  $sd = 22,8$  e  $k = 73$ ) e os restantes cursos deste departamento ( $CL = 118,8$  dias,  $sd = 18,1$  e  $k = 10$ ).

Em relação ao departamento de História (Figura 5.21), numa primeira análise concluímos que este processo é muito instável. Registaram-se tempos iniciais muito baixos, relativos ao ano de 2009, que culminaram em pontos abaixo do limite inferior de controlo. Depois, em 2011 registou-se um aumento destes tempos, resultando em pontos acima do limite superior de controlo. Em Junho de 2012, voltou a haver uma diminuição destes tempos e no início de 2013 um aumento do T proc, resultando em pontos abaixo e acima dos limites de controlo, respetivamente.

Ao estratificar o T proc pelos diferentes cursos deste departamento, também obtivemos alguns valores da estatística EWMA muito baixos, tanto para o curso de Gestão e Valorização do Património Histórico e Cultural (GVPHC) (Figura 5.21), como para o curso de Técnicas Patrimoniais e Territoriais da Indústria (História e Valorização) (TPTI) (Figura 5.22). Referimo-nos a T proc inferiores a 14 dias, ou seja, tempos para os quais a soma do T sac, T formação do júri, T reunião e T provas não excede este período. Por exemplo, os valores obtidos para o T proc do curso TPTI podem suscitar algumas dúvidas, na medida em que verificamos que o T reunião não excede 2 dias e o T provas não excede 9 dias.

Para os restantes cursos deste departamento (Figura 5.22), concluímos que o processo está sob controlo, mas obtivemos um  $C_{pku}$  muito baixo (0,20), e uma probabilidade de ter um ponto acima do limite superior de especificação igual a 0,28.

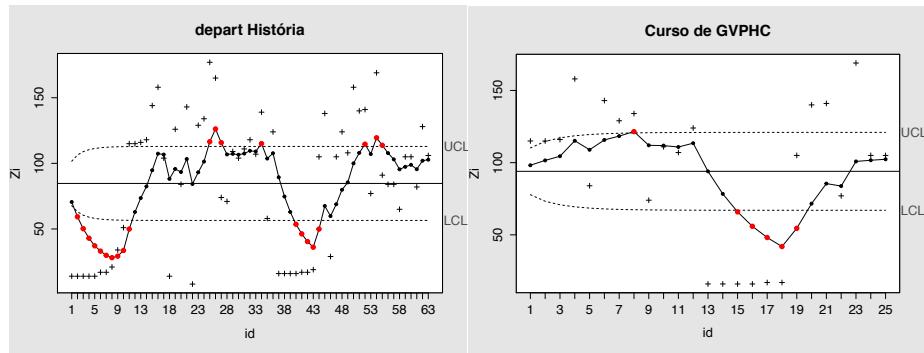


Figura 5.21: Cartas de controlo EWMA para o departamento de História ( $CL = 84,7$  dias,  $sd = 29,6$  e  $k = 63$ ) e o curso de Gestão e Valorização do Património Histórico e Cultural ( $CL = 94$  dias,  $sd = 28,3$  e  $k = 25$ ).

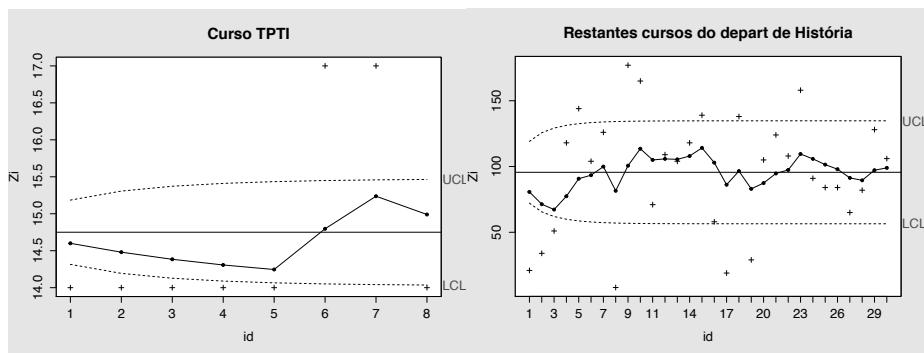
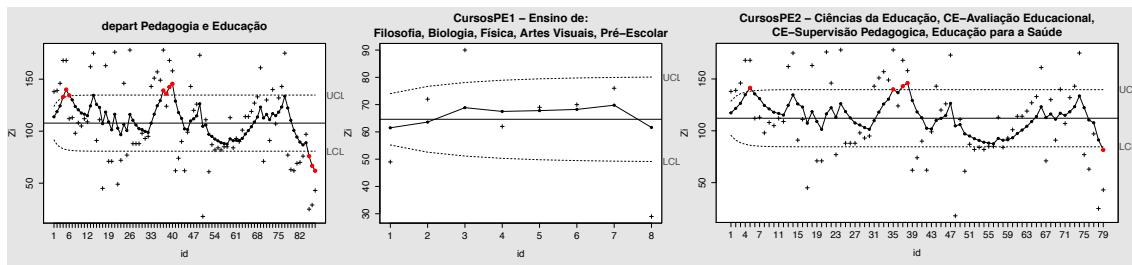


Figura 5.22: Cartas de controlo EWMA para o curso de Técnicas Patrimoniais e Territoriais da Industria (História e Valorização) ( $CL = 14,8$  dias,  $sd = 0,8$  e  $k = 8$ ) e os restantes cursos do departamento de História ( $CL = 95,6$  dias,  $sd = 41,1$  e  $k = 30$ ).

Quanto ao departamento de Pedagogia e Educação (PE), concluímos que para os cursos de Ensino o processo está sob controlo (Figura 5.23) e apresenta uma boa capacidade em cumprir as especificações superiores, na medida em que apenas 159 amostras num milhão apresentam tempos superiores aos 120 dias ( $C_{pku} = 1,12$  e Probabilidade  $< 0,0001$ ). O mesmo não se pode afirmar em relação aos cursos de Ciências da Educação, pois para estes cursos verifica-se uma maior instabilidade do processo. Neste caso, a carta reflete um aumento da média estimada no mês de Junho de 2011 e a partir de Fevereiro de 2012, culminando em pontos acima do limite de controlo. No início de Janeiro de 2013, estes tempos voltam a decrescer, resultando em um ponto abaixo do limite inferior de controlo (amostra 79).



**Figura 5.23: Cartas de controlo EWMA para o departamento de Pedagogia da Educação ( $CL = 107,7$  dias,  $sd = 28,3$  e  $k = 87$ ), os cursos de Ensino ( $CL = 64,6$  dias,  $sd = 16,5$  e  $k = 8$ ) e os cursos de Ciências da Educação ( $LC = 112,1$  dias,  $sd = 28,9$  e  $k = 79$ ).**

As cartas de controlo dos departamentos de Sociologia (Figura 5.24) e de Biologia (Figura 5.25) refletem comportamentos distintos em dois horizontes temporais (até ao ano letivo de 2010/2011 e depois desse ano letivo). Para o departamento de Sociologia houve uma redução do T proc médio estimado ( $CL_{1^{\text{a}} \text{ Período}} = 130,5$  e  $CL_{2^{\text{a}} \text{ Período}} = 86,1$ ) e para o departamento de Biologia um aumento desse tempo médio ( $CL_{1^{\text{a}} \text{ Período}} = 72,6$  e  $CL_{2^{\text{a}} \text{ Período}} = 94,2$ ). Apenas para o departamento de Sociologia, concluímos que existem diferenças significativas entre as médias estimadas dos dois períodos (valor  $p < 0,001$ ). Já para o departamento de Biologia, o valor de prova do teste de Wilcoxon indica que não existem diferenças significativas entre as médias estimadas dos dois períodos (valor  $p = 0,10$ ).

Ao investigar as causas responsáveis pelos comportamentos distintos nos diferentes períodos, concluímos que no caso do departamento de Sociologia houve um decréscimo dos T sac, enquanto que no departamento de Biologia o T formação do júri aumentou após o final do ano letivo de 2010/2011.

Importa também referir que o departamento de Sociologia apresenta uma melhor capacidade em cumprir as especificações superiores depois de 2010/2011, pois obtivemos uma probabilidade de ter um ponto acima dos 120 dias relativamente baixa (Probabilidade = 0,05 e  $C_{pku} = 0,55$ ), ao contrário do que se verificou no primeiro período (Probabilidade<sub>SN</sub> = 0,69 e  $C_{pku(SN)} = 0,07$ ). Já para o segundo período do departamento de Biologia, concluímos que este processo apresenta uma probabilidade igual a 0,19 e um  $C_{pku}$  igual a 0,29.

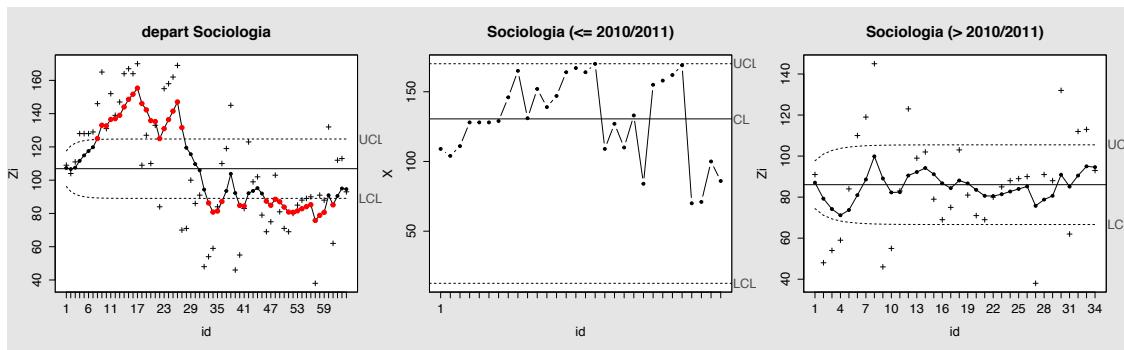


Figura 5.24: Carta EWMA para o departamento de Sociologia ( $CL = 106,9$  dias,  $sd = 18,7$  e  $k = 64$ ), Carta X-SN para o primeiro período (até ao ano letivo de 2010/2011) ( $CL = 130,5$  dias,  $sd = 17,3$ ,  $LCL = 12,6$ ,  $UCL = 170,1$  e  $k = 30$ ) e carta EWMA para o segundo período (após o ano letivo de 2010/2011) ( $CL = 86,1$  dias,  $sd = 20,4$  e  $k = 34$ ).

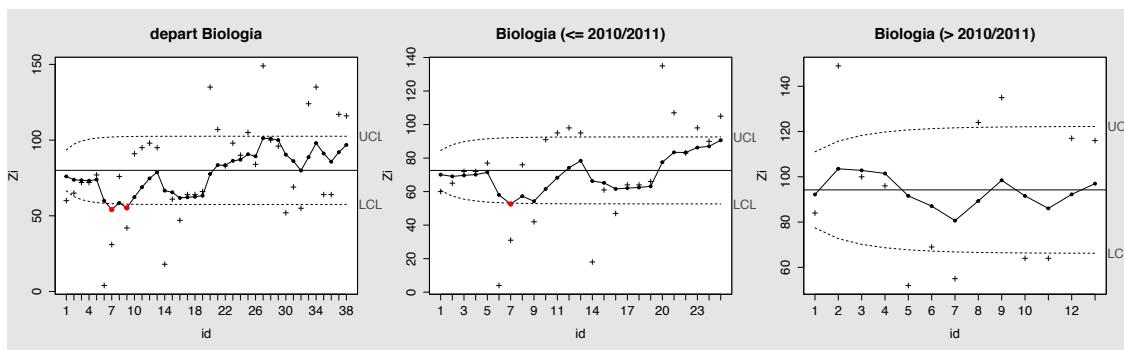


Figura 5.25: Cartas EWMA para o departamento de Biologia ( $CL = 80$  dias,  $sd = 23,6$  e  $k = 38$ ), para o primeiro período (até 2010/2011) ( $CL = 72,6$  dias,  $sd = 20,9$  e  $k = 25$ ) e para o segundo período (após 2010/2011) ( $CL = 94,2$  dias,  $sd = 29,4$  e  $k = 13$ ).

Quanto ao departamento de Desporto e Saúde (Figura 5.26), também identificamos dois períodos distintos (até ao ano letivo de 2011/2012 e depois desse ano letivo). Para a carta deste departamento, até ao final de 2011, registam-se alguns tempos acima da linha central. Ao elaborar as cartas de controlo para estes dois períodos concluímos que a média estimada do T proc até 2011/2012 ( $CL = 110,3$  dias) é superior à do período seguinte ( $CL = 88,5$  dias), porém não podemos afirmar que existam diferenças significativas entre as médias estimadas para os dois períodos (valor  $p = 0,10$ ).

Apesar de se terem registado dois períodos distintos, não conseguimos identificar causas responsáveis por este comportamento, pois neste período não se registraram mudanças significativas dos tempos parciais. Repare-se que para o segundo período deste departamento verifica-se uma maior capacidade do processo cumprir as especificações superiores ( $C_{pku} = 0,35$  e Probabilidade = 0,15) do que para o primeiro período ( $C_{pku} = 0,12$  e Probabilidade = 0,36).

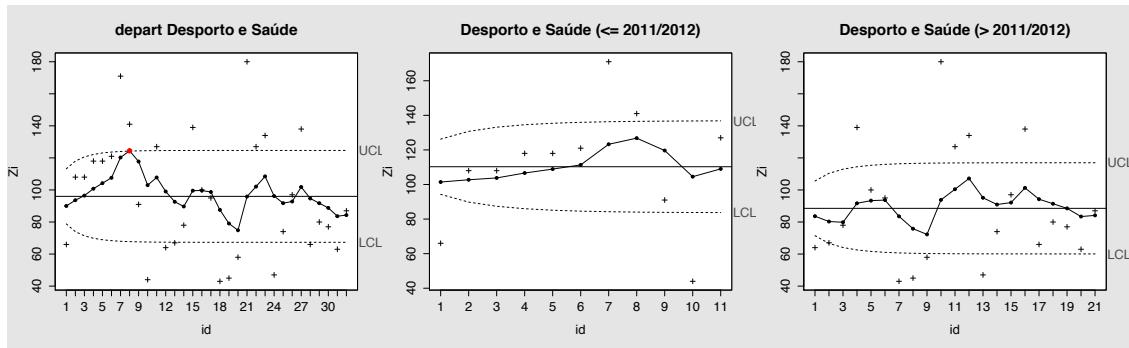


Figura 5.26: Cartas EWMA para o departamento de Desporto e Saúde ( $CL = 96$  dias,  $sd = 30,1$  e  $k = 32$ ), para o primeiro período (até 2011/2012) ( $CL = 110,3$  dias,  $sd = 27,9$  e  $k = 11$ ) e para o segundo período (após 2011/2012) ( $CL = 88,5$  dias,  $sd = 29,8$  e  $k = 21$ ).

Finalmente, apresentamos a carta relativa ao departamento de Informática (Figura 5.27). Também para este departamento registaram-se alterações ao longo do tempo. O T proc médio estimado para o primeiro período (até ao ano de conclusão de 2010) ( $CL = 71,8$  dias) é inferior ao do segundo período ( $CL = 94,6$  dias) (valor  $p = 0,01$ ). Este aumento deve-se a alterações dos processos de reunião do júri. Concluímos que nos anos posteriores a 2010 se registou um aumento considerável dos T reunião. Assim, concluímos que o processo do primeiro período apresenta uma melhor capacidade em cumprir as especificações superiores ( $C_{pku} = 1,39$ , i.e., apenas 15 amostras num milhão estão acima dos 120 dias, e Probabilidade  $< 0,001$ ) do que o processo do segundo período ( $C_{pku} = 0,25$  e Probabilidade = 0,22). Por outro lado, apesar da carta relativa ao primeiro período apresentar uma melhor capacidade em cumprir as especificações superiores, esta apresenta vários pontos acima da linha central o que indica que a média estimada pode ter aumentado naquele período. Já a carta do segundo período é uma carta que traduz um processo sob controlo, mesmo que o processo tenha uma fraca capacidade em cumprir as especificações superiores.

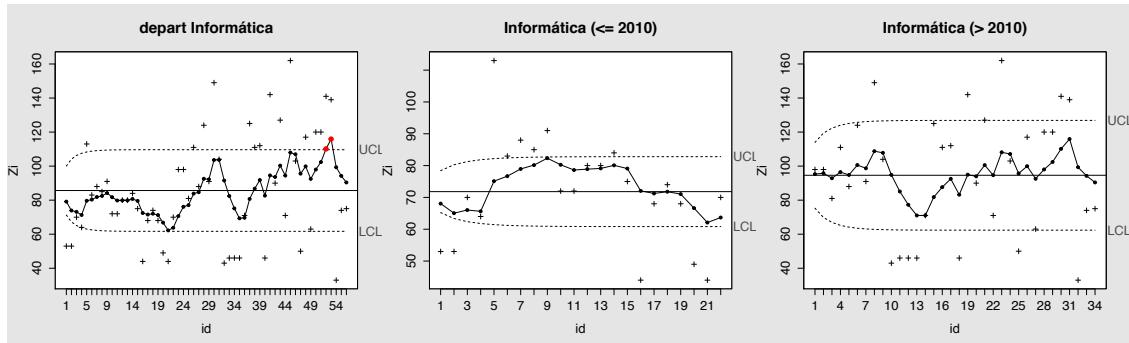


Figura 5.27: Cartas EWMA para o departamento de Informática ( $CL = 85,7$  dias,  $sd = 25,1$  e  $k = 56$ ), o primeiro período (até ao ano de conclusão de 2010) ( $CL = 71,8$  dias,  $sd = 11,5$  e  $k = 22$ ) e para o segundo período (depois do ano de conclusão de 2010) ( $CL = 94,6$  dias,  $sd = 33,8$  e  $k = 34$ ).

#### V.4.2. Tempo entre a entrega e o envio da dissertação para a escola

Segundo o regulamento interno da Universidade, no prazo máximo de 5 dias úteis após confirmação da conclusão da componente curricular, os SAC devem remeter para o Concelho científico os exemplares da dissertação. Contudo, uma análise ao T sac (Figura 5.4) permitiu concluir que para 25% desta amostra se registaram tempos entre 24 e 143 dias. Repare-se que apesar de não sabemos ao certo quando é que ocorreu a confirmação da conclusão da parte curricular, partimos do pressuposto que estes serviços não recebem as dissertações sem obter esta confirmação.

Concluímos que a distribuição destes tempos se ajusta melhor a uma distribuição SN (Tabela 2, Anexo 2), por isso optámos por construir a Carta X-SN da Figura 5.28. A partir desta carta, podemos concluir que este processo não está sob controlo, pois obtivemos alguns pontos acima do limite superior de controlo. Por outro lado, assistimos a comportamentos distintos ao longo do tempo, sendo que até Outubro de 2009 obtivemos uma média estimada inferior ( $CL = 8,3$  dias) à média estimada até ao final desse ano ( $CL = 23,8$  dias) (valor  $p = 0,03$ ) e à média de 2010 ( $CL = 28,1$  dias) (valor  $p = 0,01$ ) (Figura 5.29). Repare-se que os anos considerados nesta secção correspondem aos anos de conclusão do T sac, ou seja, aos anos em que se registou o envio da dissertação para a escola.

Os processos dos anos 2011, 2012 e 2013 são muito semelhantes (Figura 5.30), sendo que para a maioria das amostras obtivemos T sac inferiores a 20 dias e, periodicamente, registaram-se tempos muito elevados. Após investigar as causas responsáveis por este comportamento, concluímos que os tempos mais elevados correspondem às amostras dos meses de Julho e Setembro, seja qual for a escola ou departamentos considerados.

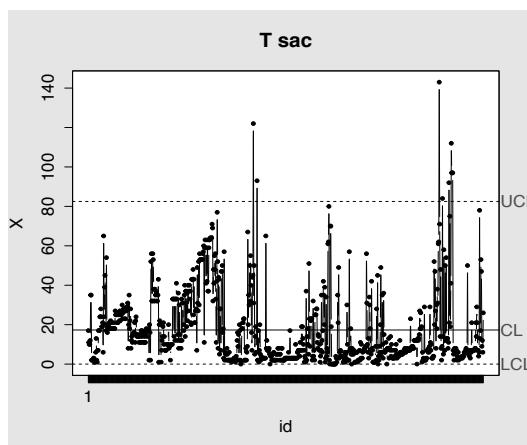


Figura 5.28: Carta X-SN para o T sac ( $CL = 17,3$  dias,  $sd = 8,5$ ,  $LCL = 0$ ,  $UCL = 82,5$  e  $k = 853$ ).

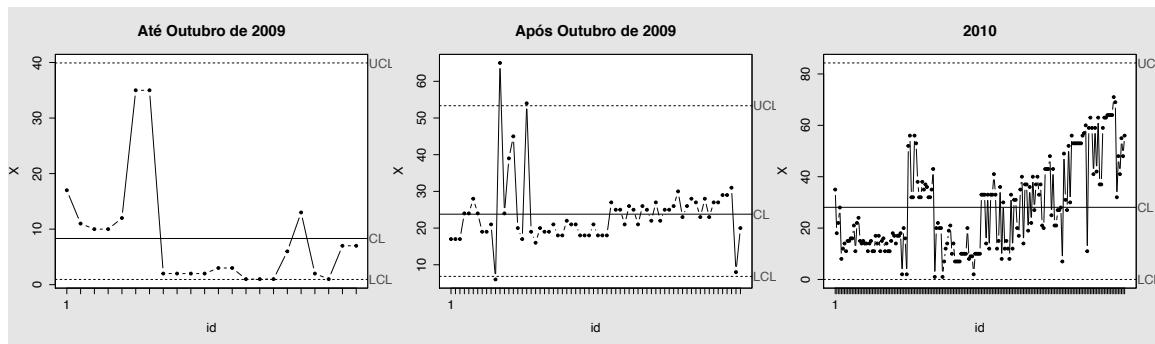


Figura 5.29: Carta X-SN para o T sac até Outubro de 2009 ( $CL = 8,3$  dias,  $sd = 4,1$ ,  $LCL = 0,94$ ,  $UCL = 39,9$  e  $k = 22$ ), para o T sac depois de Outubro de 2009 ( $CL = 23,8$  dias,  $sd = 5,6$ ,  $LCL = 6,8$ ,  $UCL = 53,3$  e  $k = 66$ ) e para o T sac de 2010 ( $CL = 28,1$  dias,  $sd = 7,8$ ,  $LCL = 0$ ,  $UCL = 84,3$  e  $k = 187$ ).

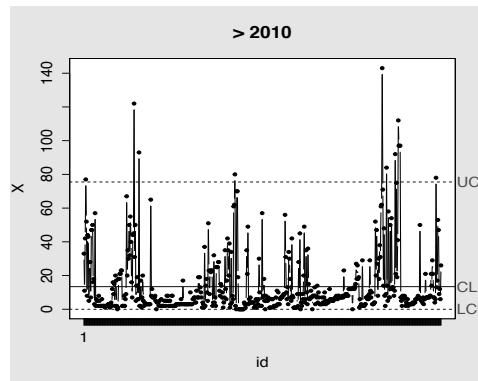


Figura 5.30: Carta X-SN para o T sac após o ano 2010 ( $CL = 13,4$  dias,  $sd = 9,2$ ,  $LCL = 0$ ,  $UCL = 75,5$  e  $k = 578$ ).

Ao estratificar o T sac pelas diferentes escolas e departamentos também não foi possível estabilizar o processo, o que apenas veio confirmar a grande instabilidade do mesmo. Repare-se que, ao considerar as cartas das diferentes escolas (Figura 5.31), os processos de CS e CT mantêm-se fora de controlo. Por outro lado, é para a Escola de CT que obtivemos uma média estimada mais elevada ( $CL = 22,6$  dias) (valor  $p < 0,001$ ). Para a Escola de Artes concluímos que existem dois pontos muito elevados, sendo o primeiro relativo a Novembro de 2009 e último relativo a Março de 2013, o que resulta numa probabilidade de obter um ponto acima dos 5 dias muito elevada (Probabilidade<sub>SN</sub> = 0,85 e  $C_{pkU(SN)} = -0,06$ ).

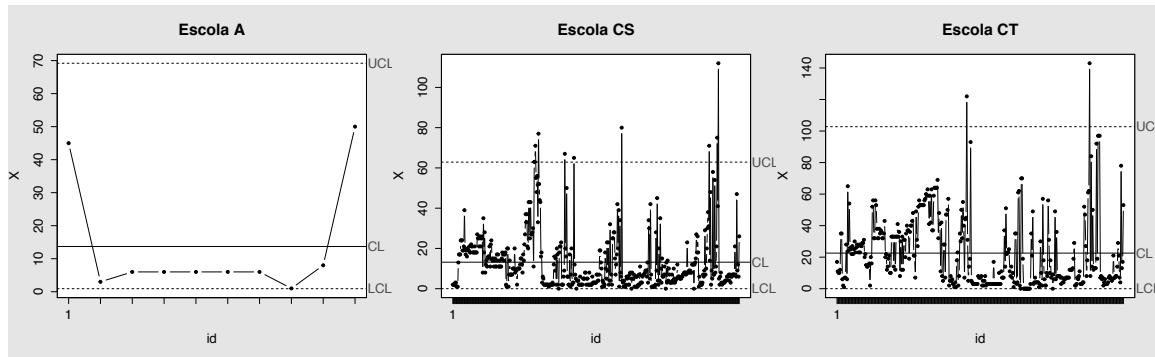


Figura 5.31: Carta X-SN para a Escola de A ( $CL = 13,7$  dias,  $sd = 9,8$ ,  $LCL = 0,9$ ,  $UCL = 69,2$  e  $k = 10$ ), para a Escola de CS ( $CL = 13,2$  dias,  $sd = 6,9$ ,  $LCL = 0$ ,  $UCL = 62,9$  e  $k = 469$ ) e para a Escola de CT ( $CL = 22,6$  dias,  $sd = 10,8$ ,  $LCL = 0$ ,  $UCL = 102,8$  e  $k = 374$ ).

Para tentar estabilizar este processo, consideramos apenas os anos posteriores a 2010 e retiramos as amostras relativas aos departamentos que apresentavam valores do T sac muito elevados (DS, GVPC, I e B), resultando numa amostra de 500 observações. Apesar de termos verificado uma redução da média estimada de 13,4 (Figura 5.30) para 10,8 dias (Figura 5.32), a tendência é sempre a mesma, pois continuamos a ter um processo muito instável. Também neste caso, é para os meses de Julho e Setembro que se verificam valores do T sac mais elevados.

Mesmo quando retiramos, sucessivamente, as amostras que estão fora dos limites de controlo, continuamos a ter pontos acima do limite superior de controlo e a questão sazonal mantém-se.

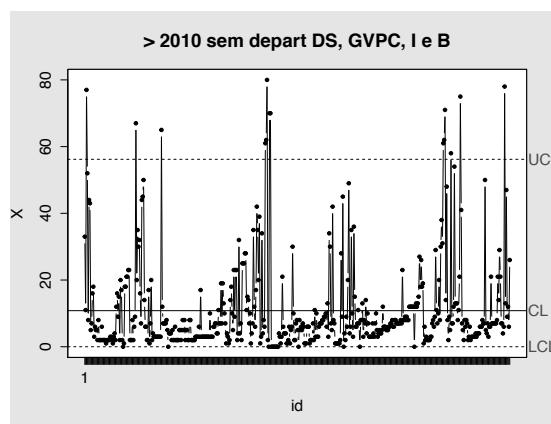


Figura 5.32: Carta X-SN para o T sac sem as amostras dos anos de 2009 e 2010 e sem dos departamentos DS, GVPC, I e B ( $CL = 10,8$  dias,  $sd = 7$   $LCL = 0$ ,  $UCL = 56,2$  e  $k = 500$ ).

#### V.4.3. Tempo de formação do júri

Segundo o regulamento interno da Universidade, depois do envio da dissertação para a escola, no prazo máximo de 30 dias consecutivos, o Diretor da Escola deverá nomear o júri. Porém uma análise ao T formação do júri (Figura 5.5) permitiu concluir que para metade desta amostra obtivemos tempos superiores a 33 dias. Por outro lado, também concluímos que o T formação do júri é responsável pela maioria dos T proc mais elevados (Figuras 5.8 a 5.12).

Para tentar compreender melhor o processo de formação do júri, recorremos a cartas de controlo de qualidade e medimos índices da capacidade do processo cumprir as especificações superiores. Tendo em conta a distribuição dos tempos em causa (Tabela 3, Anexo 2), implementamos uma Carta X-SN para o T formação do júri (Figura 5.33) e concluímos que até 2010 o comportamento deste processo distingue-se dos restantes anos. Apesar de não se verificar uma instabilidade tão acentuada como no processo dos SAC, também neste caso o processo não está sob controlo e apresenta uma média estimada superior à desejada (CL = 35,9 dias).

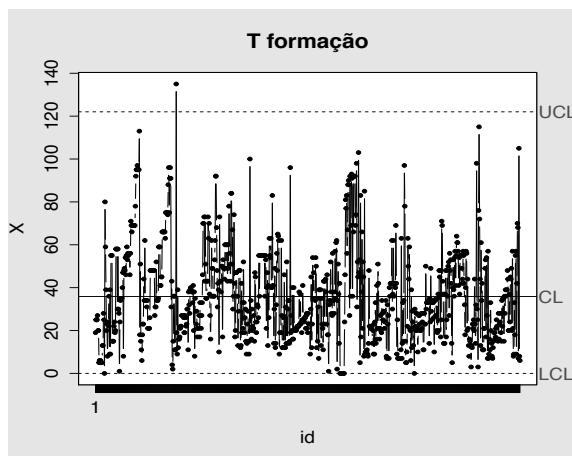


Figura 5.33: Carta X-SN para o T formação do júri (CL = 35,9 dias,  $sd = 12,5$  LCL = 0, UCL = 122 e  $k = 583$ ).

Ao analisar as cartas relativas às diferentes escolas (Figura 5.34), concluímos que é para a Escola de CS que obtivemos uma média estimada mais elevada (CL = 39,7 dias) (valor  $p < 0,001$ ). Apesar da média estimada para a Escola de A e de CT ser semelhante (aproximadamente, 31 dias), concluímos que o processo apenas está controlado para a Escola de A e que este apresenta uma probabilidade elevada de obter um ponto acima dos 30 dias (Probabilidade = 0,61 e  $C_{pku} = -0,10$ ). Para a carta da escola de CS obtivemos um ponto abaixo do limite inferior de controlo, relativo ao curso de Ciências da Educação, cujo T formação do júri é igual a zero dias. Já para a carta da escola

de CT, registou-se um ponto acima do limite superior de controlo (135 dias), que corresponde a uma observação do curso de Química em Contexto Escolar, em Junho de 2013.

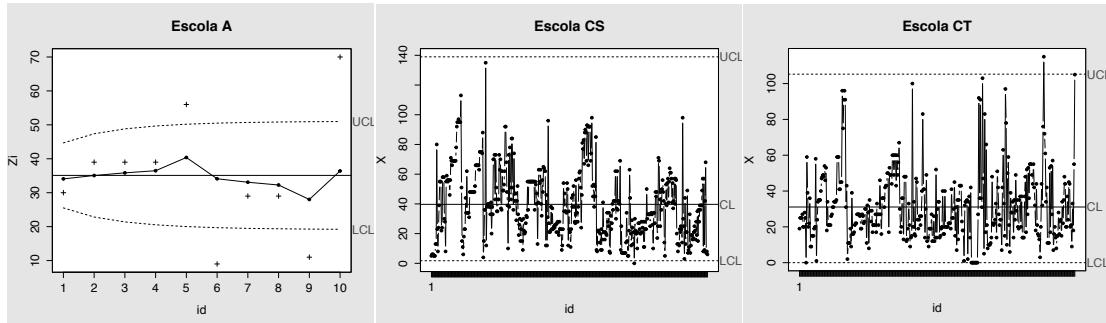


Figura 5.34: Carta EWMA para a Escola de A ( $CL = 30,5$  dias,  $sd = 16,7$  e  $k = 10$ ); Cartas X-SN para a escola CS ( $CL = 39,7$  dias,  $sd = 11,4$   $LCL = 1,74$ ,  $UCL = 139$  e  $k = 469$ ) e para a escola CT ( $CL = 31,1$  dias,  $sd = 12,7$   $LCL = 0$ ,  $UCL = 105,3$  e  $k = 374$ ).

Os anos considerados para este processo correspondem aos anos de conclusão do T formação do júri, ou seja, aos anos em que se registou a data de homologação. Também neste caso, verificamos que existem diferenças num horizonte temporal (Figura 5.35). Em 2009 a média estimada ( $CL = 26,4$  dias) é inferior à dos restantes anos (valor  $p < 0,001$ ), sendo que em 2010 obtivemos a média mais elevada ( $CL = 44,5$  dias) (valor  $p < 0,001$ ).

Apesar de não termos nenhum ponto fora dos limites de controlo para as cartas destes anos, obtivemos baixos índices da capacidade do processo cumprir as especificações superiores (30 dias). A probabilidade de obter um ponto acima dos 30 dias é elevada tanto para 2009 ( $Probabilidade_{SN} = 0,36$  e  $C_{pku(SN)} = 0,04$ ), como para 2010 ( $Probabilidade_{SN} = 0,67$  e  $C_{pku(SN)} = -0,10$ ) e depois desse ano ( $Probabilidade_{SN} = 0,51$  e  $C_{pku(SN)} = -0,04$ ).

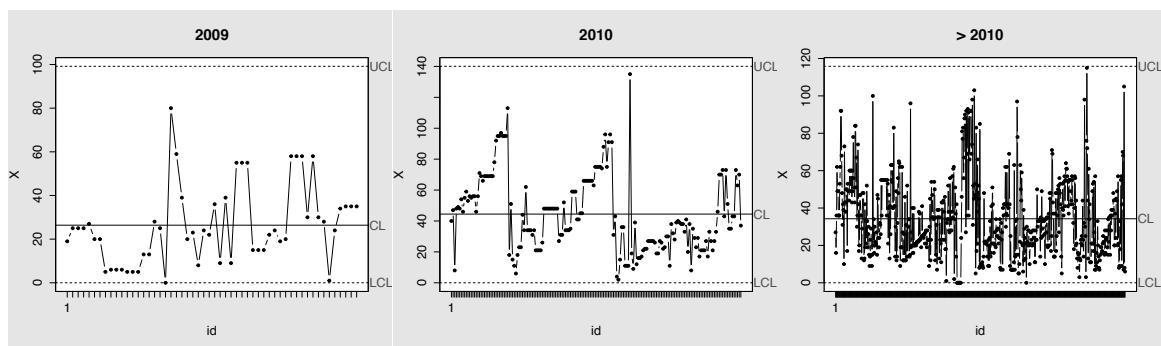


Figura 5.35: Cartas X-SN para 2009 ( $CL = 26,4$  dias,  $sd = 10,8$   $LCL = 0$ ,  $UCL = 99,2$  e  $k = 54$ ), para 2010 ( $CL = 44,5$  dias,  $sd = 9,9$   $LCL = 0$ ,  $UCL = 140,1$  e  $k = 176$ ) e para depois de 2010 ( $CL = 34,3$  dias,  $sd = 13,7$   $LCL = 0$ ,  $UCL = 115,8$  e  $k = 623$ ).

A partir de 2010, voltamos a identificar períodos para os quais estes tempos são mais elevados. Estes períodos correspondem aos meses que decorrem entre Janeiro e Março e entre Junho e Setembro, ou seja, correspondem ao final dos semestres ímpares e pares, respetivamente. Por outro lado, no início de 2009 registaram-se tempos muito baixos e os tempos mais elevados não correspondem aos períodos identificados nos restantes anos.

Ao confirmar a sazonalidade inerente a este processo, para os anos posteriores a 2009, optámos por estratificar os T formação do júri pelos diferentes períodos. Primeiro, consideramos uma carta para os meses de Janeiro a Março e de Junho a Setembro dos anos posteriores a 2009 (Figura 5.36). Depois, elaboramos uma carta para os restantes meses (Figura 5.36).

A média estimada para os tempos do final dos semestres ( $CL = 45,6$  dias) é estatisticamente superior à média estimada para os restantes meses ( $CL = 29,3$  dias) (valor  $p < 0,001$ ), o que vem confirmar a distinção entre estes dois períodos. Para os restantes meses dos diferentes anos, não existem pontos fora dos limites de controlo, porém para este processo a probabilidade de obter um ponto acima dos 30 dias é elevada (0,42) e obtivemos um  $C_{pku(SN)}$  praticamente nulo (0,007).

Quando analisamos a carta relativa ao final dos semestres, concluímos que existem muitos pontos fora dos limites inferior e superior de controlo. Repare-se que partir de 2012 (amostra 231) há uma redução destes tempos que resultam em vários pontos abaixo do limite inferior de controlo. Ao considerar apenas os tempos relativos ao final dos semestres do ano de 2012 e 2013 (Figura 5.37), concluímos que apesar de não se terem registado pontos fora dos limites de controlo, existem alguns pontos acima do limite superior de especificação e algumas observações consecutivas abaixo dos 17 dias. Este processo apresenta uma probabilidade elevada de ter um ponto acima dos 30 dias (0,49) e um  $C_{pku(SN)}$  muito baixo (0,01).

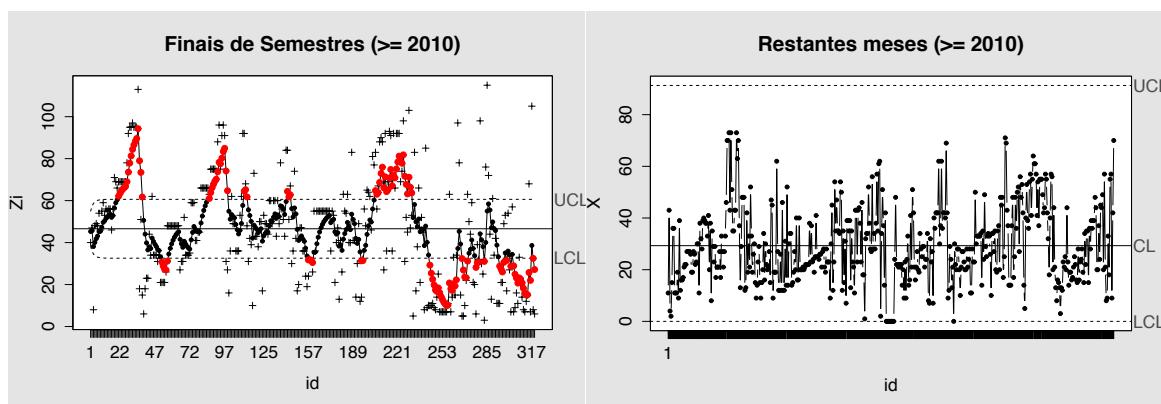


Figura 5.36: Carta EWMA para os meses de Janeiro a Março e Junho a Setembro dos anos 2010, 2011, 2012 e 2013 ( $CL = 45,6$  dias,  $sd = 14,7$  e  $k = 319$ ) e Carta X-SN para os restantes meses ( $CL = 29,3$  dias,  $sd = 10,3$   $LCL = 0$ ,  $UCL = 91,3$  e  $k = 480$ ).

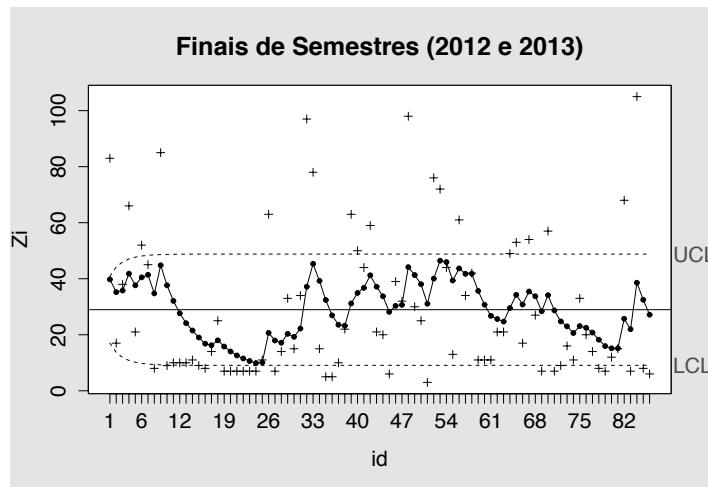


Figura 5.37: Carta EWMA para os meses de Janeiro a Março e Junho a Setembro do ano 2012 e 2013 (CL = 28,9 dias,  $sd = 20,8$  e  $k = 86$ ).

#### V.4.4. Tempo até à primeira reunião do júri

Para avaliar o processo relativo ao tempo que decorre até à primeira reunião do júri (T reunião), tivemos em consideração que o limite superior de especificação estabelecido pelo regulamento interno é igual a 30 dias. Uma análise ao T reunião permitiu concluir que esta variável segue uma distribuição SN (Tabela 4, Anexo 2). Através da Carta X-SN da Figura 5.38, verificamos que o processo relativo ao T reunião apresenta uma média estimada igual a 25,6 dias.

Ao analisar as cartas implementadas para as diferentes escolas (Figura 5.39), concluímos que a Escola de CT apresenta uma média estimada inferior à das restantes escolas (CL = 19,8 dias) (valor  $p < 0,001$ ). Tanto para a Escola de A como a Escola de CS, obtivemos baixos índices de capacidade do processo, sendo que a probabilidade de ter um ponto acima do limite superior de especificação é igual a 0,52 ( $C_{pku} = -0,02$ ) para a Escola de A e 0,42 ( $C_{pku(SN)} = -0,001$ ) para a Escola de CS.

Já quando se consideram os diferentes anos de conclusão do T reunião (anos correspondentes à data do despacho liminar), concluímos que a média estimada para o ano 2009 é inferior à dos restantes anos (CL = 8,4) (valor  $p < 0,001$ ). Nos anos seguintes, este processo não apresenta comportamentos muito distintos e as médias estimadas são semelhantes (Figuras 5.40 e 5.41). Por outro lado, verificamos que o processo não apresenta pontos fora dos limites de controlo para as cartas de 2009, 2010 e 2013. Em 2009 obtivemos melhores índices da capacidade do processo cumprir as especificações superiores, sendo que a probabilidade de ter um ponto acima

dos 30 dias é igual a 0,01 em 2009 ( $C_{pku(SN)} = 0,60$ ), 0,38 em 2010 ( $C_{pku(SN)} = 0,04$ ) e 0,40 em 2013 ( $C_{pku(SN)} = 0,01$ ).

Ao analisar as cartas relativas aos diferentes anos de conclusão, verificamos que também para este processo parece haver uma tendência para obter tempos mais elevados em períodos distintos. Efetivamente, concluímos que é para os meses de Fevereiro, Março e Outubro que se registam T reunião superiores.

Para implementar uma carta de controlo para este processo, optamos por retirar as amostras cujos T reunião estavam fora do limite superior de controlo, o que resultou na eliminação das amostras com tempos superiores a 100 dias. Na Figura 5.42 apresentamos a carta de controlo para este processo. Apesar de não obtermos nenhum ponto fora dos limites de controlo, este processo não tem uma boa capacidade de cumprir as especificações superiores (Probabilidade<sub>SN</sub> = 0,35 e  $C_{pku(SN)} = 0,05$ ).

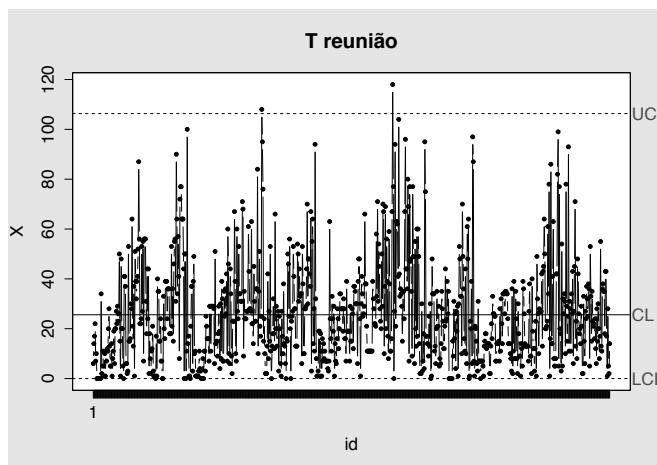


Figura 5.38: Carta X-SN para o T reunião (CL = 25,6 dias,  $sd = 14,5$  LCL = 0, UCL = 106,2 e  $k = 583$ ).

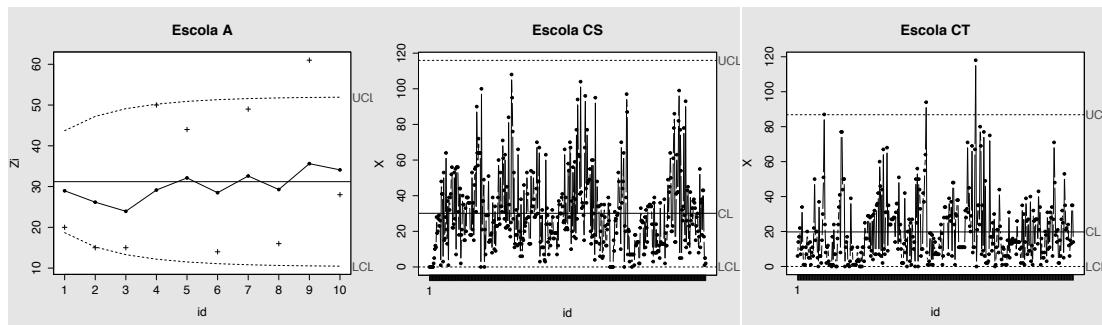


Figura 5.39: Carta EWMA para a Escola de A (CL = 31,2 dias,  $sd = 21,9$  e  $k = 10$ ); Cartas X-SN para a escola CS (CL = 30,1 dias,  $sd = 14,8$ , LCL = 0, UCL = 115,9 e  $k = 469$ ) e para a escola CT (CL = 19,8 dias,  $sd = 13,4$  LCL = 0, UCL = 86,8 e  $k = 374$ ).

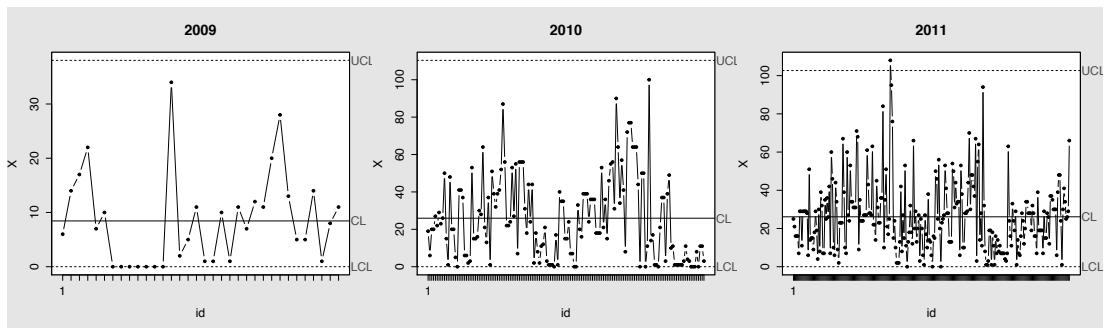


Figura 5.40: Cartas X-SN para 2009 ( $CL = 8,4$  dias,  $sd = 6,4$  LCL = 0, UCL = 38 e  $k = 34$ ), para 2010 ( $CL = 25,9$  dias,  $sd = 14,2$  LCL = 0, UCL = 110,4 e  $k = 152$ ) e para 2011 ( $CL = 26,1$  dias,  $sd = 13,7$  LCL = 0, UCL = 102,7 e  $k = 263$ ).

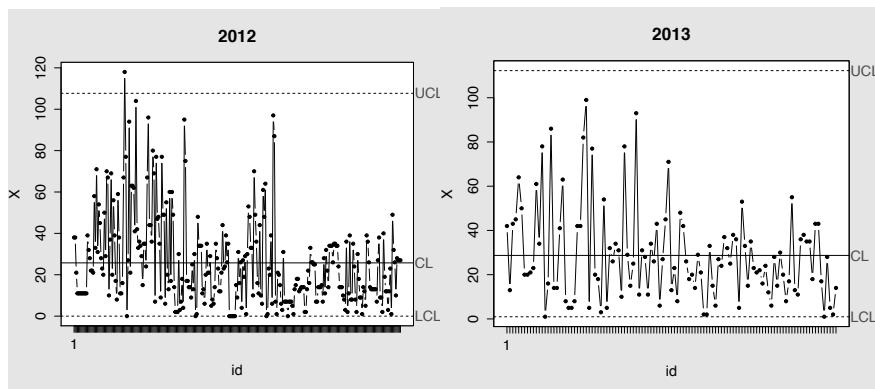


Figura 5.41: Cartas X-SN para 2012 ( $CL = 25,7$  dias,  $sd = 14,4$  LCL = 0, UCL = 107,7 e  $k = 291$ ) e para 2013 ( $CL = 28,7$  dias,  $sd = 19,1$  LCL = 0,96, UCL = 112,2 e  $k = 113$ ).

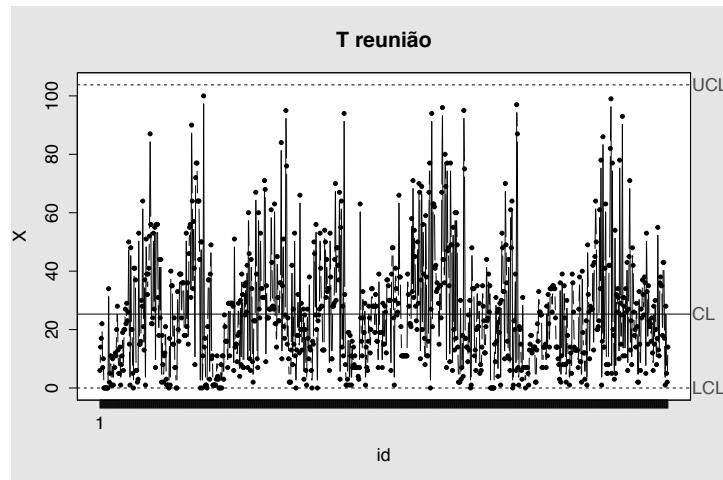


Figura 5.42: Cartas X-SN para o T reunião sem as amostras com tempos superiores a 100 dias ( $CL = 25,3$  dias,  $sd = 14,3$  LCL = 0, UCL = 103,8 e  $k = 850$ ).

#### V.4.5. Tempo entre a primeira reunião e a realização das provas

Segundo o regulamento interno, o tempo que decorre entre a primeira reunião do júri e a realização das provas (T provas) não pode exceder os 60 dias. Para compreender melhor este processo, construímos uma Carta X-SN, mais adequada à distribuição dos tempos em causa (Tabela 5, Anexo 2). Ao implementar a carta da Figura 5.43, concluímos que o processo apresenta uma probabilidade de ter um ponto acima do limite superior de especificação (60 dias) igual a 0,06. Por outro lado, obtivemos um  $C_{pku}$  baixo ( $C_{pku(SN)} = 0,32$ ), na medida em que 165662 amostras num milhão estão acima dos 60 dias.

Apesar do processo estar sob controlo, optamos por estratificar as amostras pelas diferentes escolas e anos de conclusão (anos correspondentes à data das provas), para de alguma forma tentar compreender se este processo se distingue em horizontes temporais. Contudo, não identificamos períodos para os quais estes tempos são mais elevados, nem podemos considerar que haja um comportamento distinto entre as escolas de CS e CT (Figura 5.44) e entre os anos posteriores a 2009 (Figuras 5.45 e 5.46). À semelhança do que se verificou no processo dos restantes tempos parciais, o processo apresenta uma média estimada mais elevada no caso da Escola de A ( $CL = 35,7$  dias, Probabilidade = 0,24 e  $C_{pku} = 0,24$ ) e uma média mais baixa para o ano de 2009 ( $CL = 10,6$  dias, Probabilidade<sub>SN</sub> < 0,01 e  $C_{pku(SN)} = 1,19$ ).

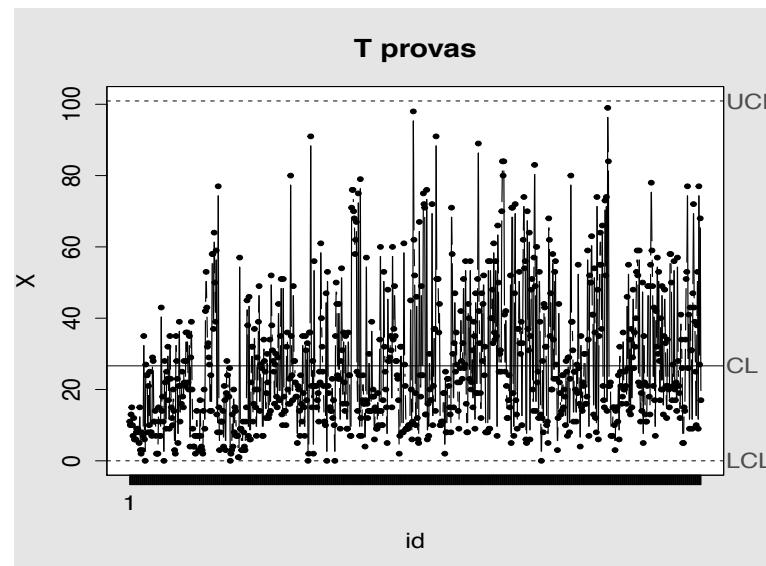


Figura 5.43: Carta X-SN para o T provas ( $CL = 26,6$  dias,  $sd = 14,1$  LCL = 0, UCL = 100,9 e  $k = 583$ ).

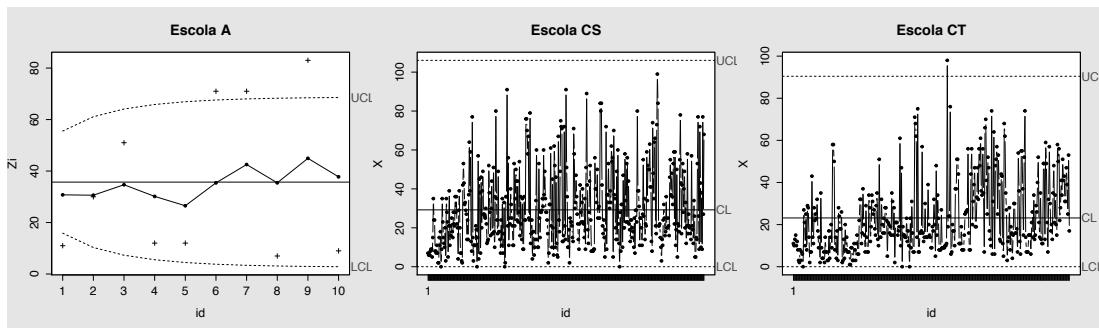


Figura 5.44: Carta EWMA para a Escola de A ( $CL = 35,7$  dias,  $sd = 34,7$  e  $k = 10$ ); Cartas X-SN para a escola CS ( $CL = 29,2$  dias,  $sd = 15,5$ ,  $LCL = 0$ ,  $UCL = 106,1$  e  $k = 469$ ) e para a escola CT ( $CL = 23,2$  dias,  $sd = 12,5$   $LCL = 0$ ,  $UCL = 90,4$  e  $k = 374$ ).

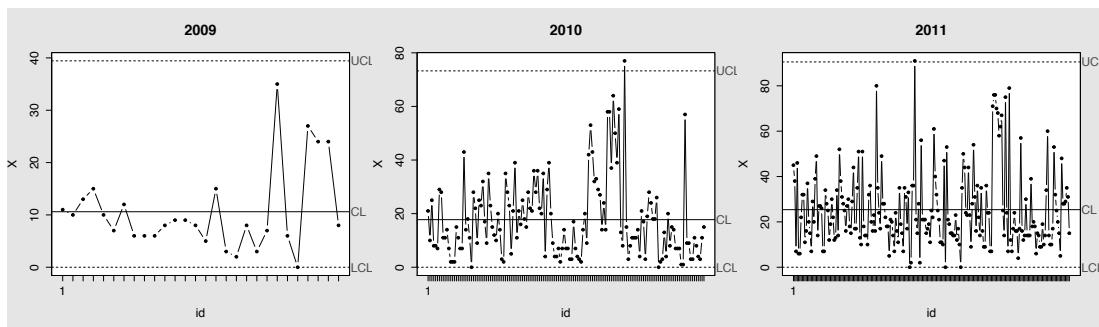


Figura 5.45: Cartas X-SN para 2009 ( $CL = 10,6$  dias,  $sd = 5,9$   $LCL = 0$ ,  $UCL = 39,4$  e  $k = 28$ ), para 2010 ( $CL = 17,8$  dias,  $sd = 9,6$ ,  $LCL = 0$ ,  $UCL = 73,3$  e  $k = 147$ ) e para 2011 ( $CL = 56,4$  dias,  $sd = 13,2$   $LCL = 0$ ,  $UCL = 90,5$  e  $k = 217$ ).

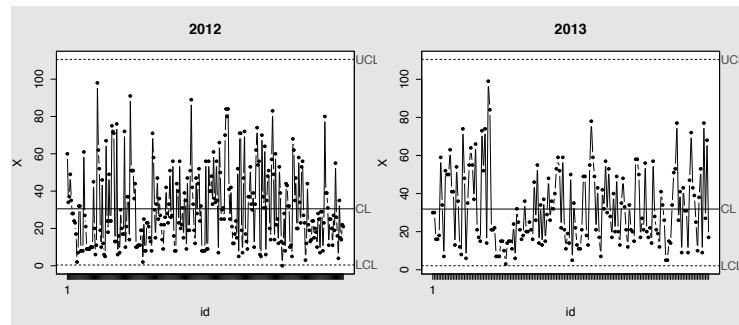


Figura 5.46: Cartas X-SN para 2012 ( $CL = 30,5$  dias,  $sd = 16,1$   $LCL = 0,4$ ,  $UCL = 110,6$  e  $k = 286$ ) e para 2013 ( $CL = 31,9$  dias,  $sd = 16,9$   $LCL = 2,1$ ,  $UCL = 110,5$  e  $k = 175$ ).

#### V.4.6. Tempo de espera pela emissão de diplomas

Para avaliar o processo relativo ao tempo de espera pela emissão de um diploma de grau (TD grau) e de curso (TD curso), tivemos em consideração que o limite superior de especificação estabelecido pelo regulamento interno é igual a 30 dias.

Uma análise ao TD grau inferior ou igual a 40 dias permitiu concluir que esta variável segue uma distribuição SN (Tabela 6, Anexo 2). Através da Carta X-SN da Figura 5.47, verificamos que o processo relativo ao TD grau apresenta uma média estimada igual a 7,3 dias.

Ao estratificar o TD grau pelos anos de conclusão (anos relativos à data de emissão do diploma) (Figuras 5.48 e 5.49), concluímos que apenas para o ano de 2010 é que o processo está controlado. Contudo, para este processo, verificamos que a média estimada para o tempo de espera pela emissão de um diploma de grau ( $CL = 11,5$  dias) é superior à do processo em que consideramos todos os anos. Para este ano, a probabilidade de ter uma amostra acima do limite superior de especificação (30 dias) é igual a 0,02, contudo obtivemos um  $C_{pku(SN)}$  muito baixo (0,45).

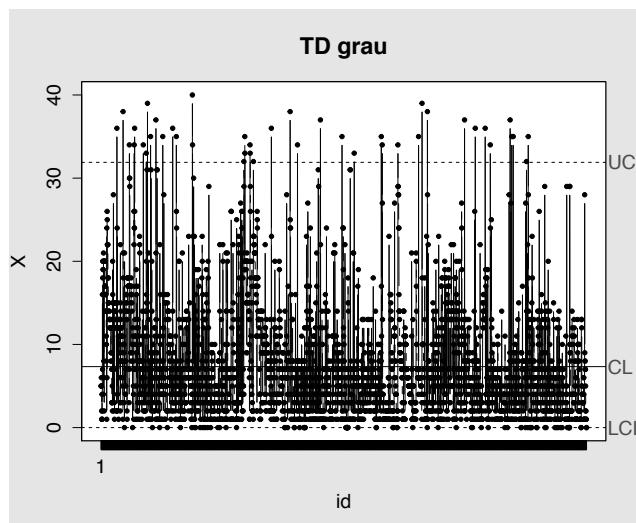


Figura 5.47: Carta X-SN para TD grau ( $CL = 7,3$  dias,  $sd = 2,8$  LCL = 0, UCL = 31,9 e  $k = 3622$ ).

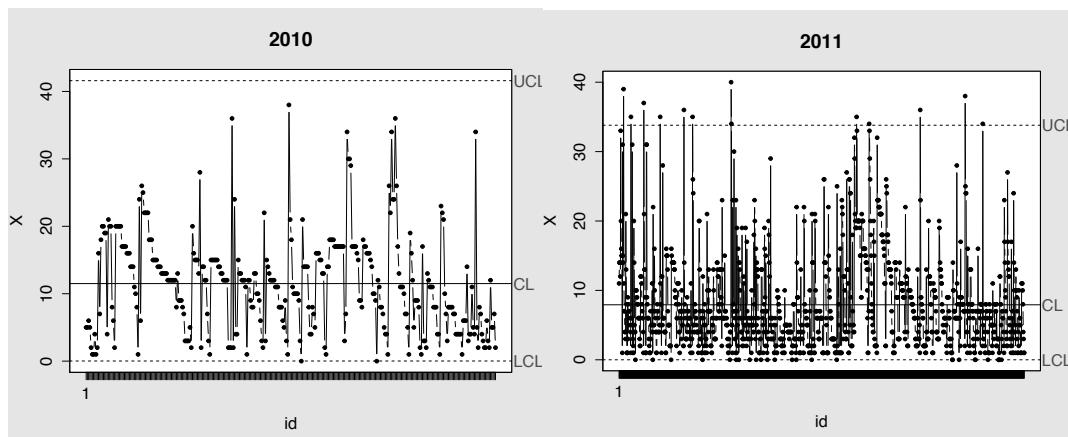


Figura 5.48: Carta X-SN para TD grau de 2010 ( $CL = 11,5$  dias,  $sd = 3,4$  LCL = 0,02, UCL = 41,6 e  $k = 332$ ) e 2011 ( $CL = 7,9$  dias,  $sd = 3,1$ , LCL = 0, UCL = 33,8 e  $k = 1263$ ).

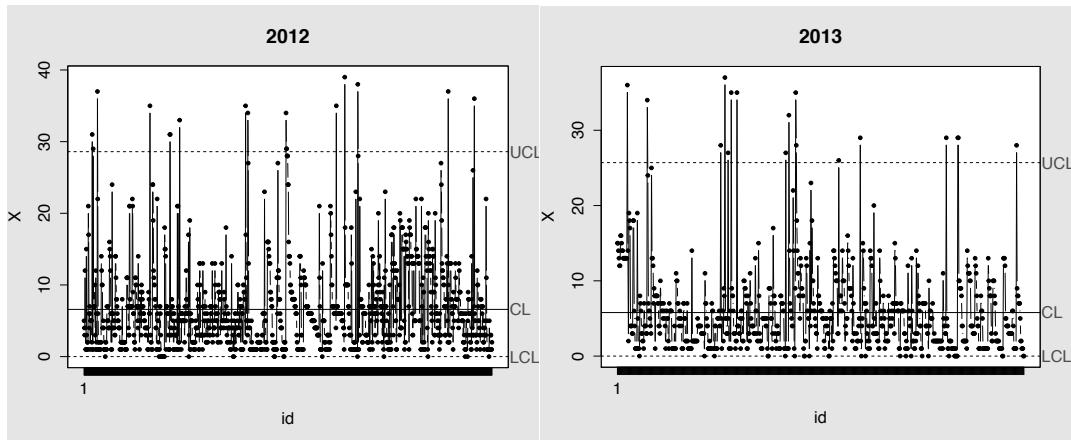


Figura 5.49: Carta X-SN para TD grau de 2012 (CL = 6,6 dias,  $sd = 2,6$ , LCL = 0, UCL = 28,6 e  $k = 1251$ ) e 2013 (CL = 5,8 dias,  $sd = 2,6$ , LCL = 0, UCL = 25,7 e  $k = 776$ ).

Ao analisar o TD curso inferior ou igual a 40 dias concluímos que também esta variável segue uma distribuição SN (Tabela 6, Anexo 2). Apesar da média estimada para o processo do TD curso ser um pouco inferior (CL = 6,3 dias) à do processo do TD grau, concluímos que este processo também não está sob controlo (Figura 5.50).

Ao estratificar o TD curso pelos anos de conclusão (Figuras 5.51 e 5.52), voltamos a concluir que apenas para o ano de 2010 é que o processo está sob controlo. Para as cartas dos restantes anos, retirámos as observações que estavam acima dos limites de controlo superiores, porém não foi possível estabilizar nenhum dos processos.

Repare-se que para o ano de 2010 a distribuição dos tempos em causa aproxima-se de uma distribuição Normal (Tabela 6, Anexo 2), logo recorremos a uma Carta EWMA e a uma Carta X de Shewhart (Figura 5.51). A média estimada para este ano é igual a 8,2 dias e, apesar de em ambas as cartas verificarmos um aumento do TD curso em Dezembro de 2010 (amostra 26), concluímos que este processo apresenta bons índices de cumprir as especificações superiores. Obtivemos uma probabilidade de uma amostra exceder os 30 dias muito baixa (Probabilidade<sub>SN</sub> < 0,001) e concluímos que não chegamos a ter 1 amostra num milhão acima do limite superior de especificação ( $C_{pku(SN)} = 1,68$ ).

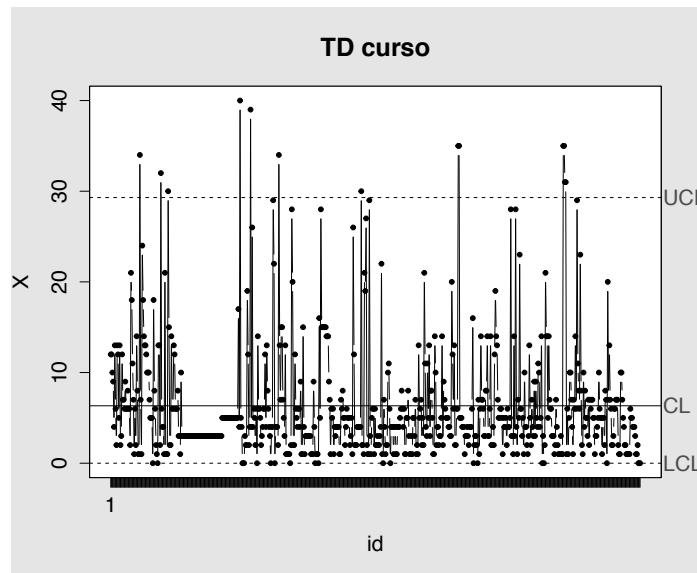


Figura 5.50: Carta X-SN para TD curso (CL = 6,3 dias,  $sd = 4$  LCL = 0, UCL = 29,3 e  $k = 656$ ).

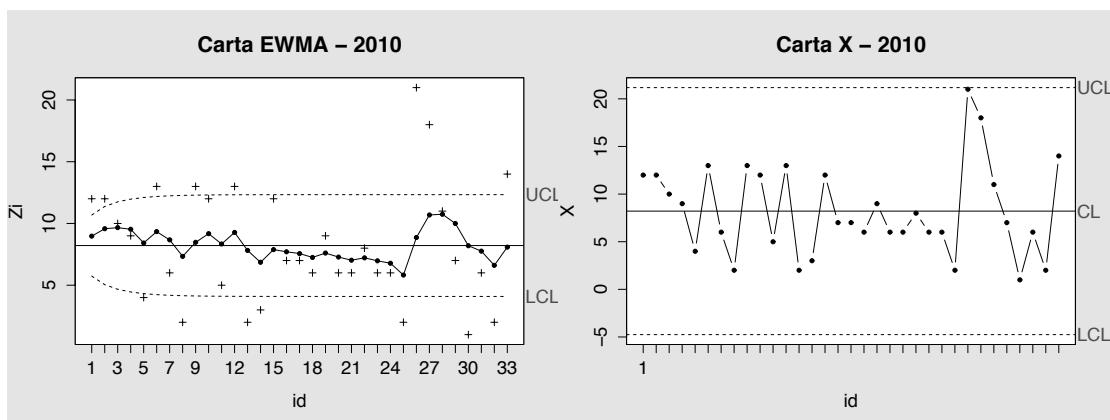


Figura 5.51: Carta EWMA (CL = 8,2 dias,  $sd = 4,3$  e  $k = 33$ ) e Carta X de Shewhart (CL = 8,2 dias,  $sd = 4,3$ , LCL = 0, UCL = 21,2 e  $k = 33$ ) para TD curso de 2010.

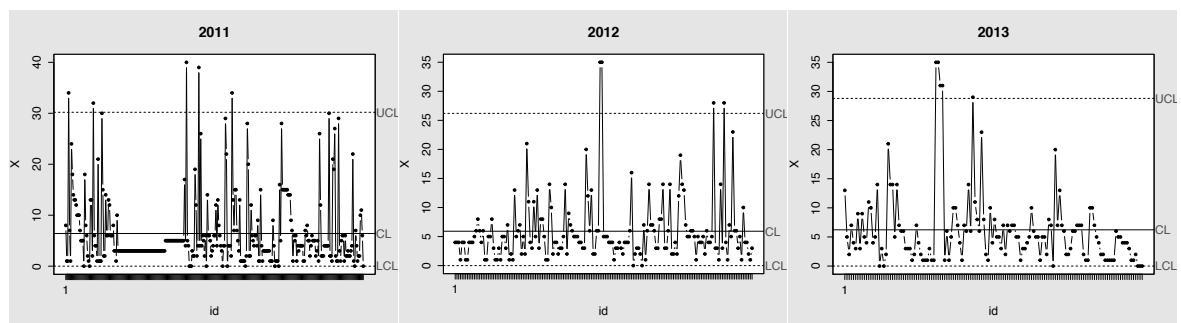


Figura 5.52: Cartas X-SN para 2011 (CL = 6,4 dias,  $sd = 4,2$ , LCL = 0, UCL = 30,2 e  $k = 314$ ), 2012 (CL = 5,9 dias,  $sd = 3,8$ , LCL = 0, UCL = 26,2 e  $k = 171$ ) e 2013 (CL = 6,2 dias,  $sd = 3,5$ , LCL = 0, UCL = 28,8 e  $k = 138$ ).

## V.5. Conclusões

Na generalidade dos processos analisados, concluímos que estes estão longe de cumprir as especificações superiores exigidas pelo regulamento interno da Universidade. Ao estratificar alguns processos pelas diferentes escolas e anos de conclusão, concluímos que os processos da Escola de A e do ano de 2009 estão quase sempre sob controlo. Porém, nem sempre apresentam uma boa capacidade em cumprir as especificações superiores.

Este estudo também permitiu concluir que estes processos apresentam uma grande instabilidade e, para alguns casos, pudemos identificar períodos em que o comportamento dos processos se distingue. Nesta situação estão os processos do T sac e do T formação do júri, que após 2010 apresentam comportamentos distintos dos anos anteriores. O mesmo se pode concluir em relação a alguns cursos do processo do T proc, nomeadamente os cursos de Psicologia e dos departamentos de Sociologia, Desporto e Saúde, Biologia e Informática, que no final de alguns anos letivos aumentaram ou diminuíram a média estimada dos respetivos processos.

Ao analisar o processo do T proc, T sac, T formação do júri e T reunião, concluímos que existe uma sazonalidade inerente a estes processos e que no final dos semestres pares e ímpares há uma maior tendência para obter tempos superiores aos exigidos pelo regulamento interno da Universidade. Também concluímos que os processos relacionados com os SAC (T sac) e com a emissão de diplomas (TD grau e TD curso) apresentam mais pontos fora de controlo e, pela abordagem usual, não foi possível colocar as respetivas cartas sob controlo.

Importa também destacar a importância da distribuição dos tempos em estudo, pois as conclusões que podemos retirar ao implementar uma Carta EWMA ou uma Carta X-SN podem ser muito distintas. Desta forma, a escolha das cartas a implementar deve ser a mais adequada para que se possa distinguir corretamente entre a variação natural dos processos e outras formas de variação, responsáveis por uma melhor ou pior prestação dos serviços em causa.

Adicionalmente, concluímos que a implementação deste tipo de ferramentas, permite dar um *feedback* aos serviços desta Universidade de modo a corrigir as causas de variação que não são inerentes a estes processos. Eventualmente, também será importante alterar o regulamento interno da Universidade, pois há casos em que a variabilidade inerente aos processos não permite que se atinjam elevados níveis de cumprimento das especificações. Assim sendo, a qualidade destes serviços dificilmente atingirá níveis de excelência.



## Capítulo VI – Considerações finais

O estudo efetuado apresentou algumas limitações. Na análise do tempo até à entrega de uma dissertação de Doutoramento recorremos a uma amostra demasiado pequena e não tivemos informação de outras variáveis que, segundo alguns autores, se adequam melhor a este grau. Nomeadamente, variáveis relacionadas com o financiamento dos alunos e com o seu grau de satisfação em relação ao apoio dos diferentes departamentos e orientadores, que podem ser preditoras do tempo até à entrega de uma dissertação.

Adicionalmente, a base de dados não continha o registo de todos os alunos e, em alguns casos, estes registos não estavam corretos, pois obtivemos tempos não credíveis, tanto para o tempo até à entrega de uma dissertação como para os diferentes tempos de procedimentos. Esta falta de informação culminou na eliminação de alguns alunos da base de dados.

Apesar das dificuldades encontradas, esta dissertação permitiu identificar alguns fatores que têm uma influência direta no tempo que decorre entre a aprovação do projeto de dissertação e a realização de provas públicas. De uma forma geral, concluímos que o adiamento da conclusão de uma dissertação de Mestrado está relacionado com a prestação dos alunos ao longo do curso e com os processos que decorrem após a entrega da dissertação nos SAC.

Concluímos que não são as características sociodemográficas do aluno nem o seu *background* familiar que determinam o adiamento da sua passagem pela Universidade. Porém, a sua idade, o ano de ingresso, a fase de candidatura, a duração da parte curricular e o facto de terem obtido o estatuto de trabalhador estudante, aquando do ingresso na Universidade, são determinantes para explicar este adiamento. Efetivamente, este estudo permitiu concluir que os alunos que ingressaram na Universidade depois da segunda fase de candidatura, nos anos de 2007 ou 2008, que não terminaram a parte curricular no tempo previsto e que beneficiaram do estatuto de trabalhador estudante quando ainda jovens têm um risco muito elevado de demorarem mais tempo a terminar a sua dissertação de Mestrado.

Por outro lado, para os cursos de Doutoramento, através de uma análise não paramétrica, concluímos que os alunos de nacionalidade Portuguesa, que residem em áreas predominantemente urbanas, com ingresso na Universidade em 2007 ou 2008 e que não terminaram a parte curricular no tempo previsto apresentam uma maior probabilidade de terminar o curso mais tarde.

Também pudemos identificar o perfil mais provável do aluno que efetua o pedido de pelo menos um semestre adicional para terminar um curso de Mestrado. Concluímos que são os alunos

de nacionalidade portuguesa, que ingressaram em 2007 ou 2008, na Escola de CS ou A e que obtiveram mais do que 25% de ECTS que apresentam uma maior probabilidade de pedir pelo menos um semestre adicional. Ainda concluímos que quanto maior é a idade de ingresso, maiores são as probabilidades destes alunos efetuarem este pedido. Por outro lado, concluímos que quanto mais recente é o ano de ingresso, menores são as probabilidades destes alunos pedirem semestres adicionais.

A identificação das causas de variação que estão na origem do adiamento da conclusão de alguns cursos pode permitir à Universidade implementar políticas de intervenção com o objetivo de prevenir os estudantes com comportamentos desfavoráveis no ensino superior. Efetivamente, se for possível perceber porque é que os alunos das escolas de A ou CS apresentam maiores probabilidades de efetuar o pedido de pelo menos um semestre adicional, será possível implementar medidas que venham suavizar as diferenças verificadas. Por outro lado, a Universidade deverá implementar medidas que venham favorecer aqueles alunos com o estatuto de trabalhador estudante ou aqueles que ingressaram na Universidade depois da segunda fase, pois também estes apresentam um maior risco de terminar a sua dissertação mais tarde.

Em relação aos processos da própria Universidade, concluímos que estes apresentam uma grande instabilidade e que estão longe de cumprir as especificações superiores estabelecidas pelo regulamento interno, sejam estas especificações relativas ao tempo de procedimentos global, ao tempo entre a entrega da dissertação e o seu envio para a escola, ao tempo de formação ou ao tempo até à primeira reunião do júri, ao tempo até à realização das provas públicas ou ao tempo de espera pela emissão de diplomas.

Apesar dos processos relativos à emissão de diplomas de curso terem cumprido as especificações estabelecidas pelo regulamento interno no ano de 2010, de uma forma geral, a Universidade deve tentar reduzir a variabilidade destes processos, cuja responsabilidade é dividida entre os SAC, os diretores de curso, a comissão de curso e o júri nomeado. Para tal, é necessário implementar medidas que permitam trazer o processo para os valores especificados. Estas medidas passam por atenuar as diferenças que existem entre os processos das diferentes escolas e entre alguns períodos temporais, nomeadamente entre os meses relativos ao final dos semestres e os restantes meses.

Adicionalmente, é preciso perceber qual a razão de se ter registado alguns valores dos tempo de procedimentos globais extremamente baixos. Para alguns dos cursos analisados, os tempos de procedimentos globais não excederam os 14 dias, o que indica que a soma dos restantes

tempos parciais não excedeu este período. Efetivamente, não conseguimos explicar a razão pela qual alguns dos tempos de formação do júri, reunião do júri e discussão de provas públicas não excederam 1 dia.

Devido à sazonalidade inerente aos processos administrativos desta Universidade, não foi possível proceder à monitorização dos mesmos. Poderíamos introduzir uma componente sazonal de forma a poder monitorizar estes processos, contudo esta abordagem está fora do âmbito desta tese. Neste sentido, esta poderá ser uma futura abordagem para tentar monitorizar este tipo de processos.

Para além do prolongamento da passagem dos alunos pelas Universidades Portuguesas, nos últimos anos temos assistido a um aumento do número de alunos que não terminam o seu curso superior. Tendo em conta que este também é um fator com um grande impacto nos orçamentos das Universidades, em futuros estudos, seria pertinente analisar o comportamento dos alunos numa perspetiva semelhante à de alguns autores. Neste sentido, seria de particular interesse estudar os fatores que levam os alunos a desistir de um curso de Mestrado ou de Doutoramento.



## Capítulo VII – Referências

- Adebi, J., & Benkin, E. (1987). The Effects of Students Academic, Financial and Demographic Variables on Time to Doctorate. *Research in Higher Education*, 27 (1), pp. 3-14.
- Azzalini, A. (2005). The Skew-Normal Distribution and Related Multivariate Families. *Scandinavian Journal of Statistics*, 32, 159-188.
- Figueiredo, F., Gomes M. (2013). The Skew-Normal Distribution in SPC. *REVSTAT – Statistical Journal*, 11 (1), 83-104.
- Green Jr., K., Toms, L., Stinson, T. (2012). Statistical Process Control applied within an Educational Services Environment. *Academy of Educational Leadership Journal*, 16 (2), 33-46.
- Gury, N. (2011). Dropping out of higher education in France: a micro-economic approach using survival analysis. *Education Economics*, 19:1, 51-64, DOI: 10.1080/09645290902796357
- HEFCE (2005). PhD Research Degrees: Entry and Completion. *Higher Education Funding Council for England*. Londres. Disponível em: <http://webarchive.nationalarchives.gov.uk/>
- Hosmer, D., Lemeshow, S. (2013). *Applied Logistic Regression* (3rd Edition). John Wiley & Sons, New York.
- Hosmer, D., Lemeshow, S. (2008). *Applied Survival Analysis* (3rd Edition). John Wiley & Sons, New York.
- Hosoi, S. A., Canetto, S. S. (2011). Women in graduate engineering: is differential dropout a factor in their under representation among engineering doctorates? *Journal of Women and Minorities in Science and Engineering* 17(1), 11-27.
- Jones-White, D., Glushko, O. (2013). The impact of the student experience on time to the Doctorate. *Office of Institutional Research*. Universidade do Minnesota, EUA. Disponível em: [www.oir.umn.edu/papers\\_and\\_publications](http://www.oir.umn.edu/papers_and_publications)
- Júnior, P., Silveira, F., Ostermann, F. (2012). Análise de sobrevivência aplicada ao estudo do fluxo escolar nos cursos de graduação em Física: um exemplo de uma universidade brasileira. *Revista Brasileira de Ensino de Física*, 34 (1), 1403.1-1403.10.
- Kleinbaum, D., Klein, M. (2012). *Survival Analysis* (3rd Edition). Springer, New York.

Min, Y., Zhang, G., Long, R., Anderson, T., Ohland, M. (2011). Nonparametric Survival Analysis of the Loss Rate of Undergraduate Engineering Students. *Journal of Engineering Education*, Vol. 100, No. 2, pp. 349-373.

Montgomery, D. C. (2009). *Introduction to Statistical Quality Control* (6th Edition). John Wiley & Sons, New York.

Owen, D.B. (1956). Tables for computing bivariate normal probabilities, *Ann. Math. Statist.*, 27, 1075–1090.

Quinn, J. (2013). Drop out and Completion in Higher Education in Europe among students from under-represented groups. *Network of Experts in social aspects of Education and Training* (NESET). Reino Unido. Disponível em: <http://www.nesetweb.eu/>

Roberts, S. W. (1959): Control Chart Tests Based on Geometric Moving Averages. *Technometrics* 1, 239–250.

Siegfried, J., Stock, W. (1999). The Labor Market for New PhD Economists. *Journal of Economic Perspectives*. V. 13, No. 3, pp. 115-134.

Teixeira, V., Infante, P., Dias, J. (2012). Modelação do tempo até conclusão de um curso superior. *Estatística Novos Desenvolvimentos e Inspirações*. (Eds. Maia, M., Campos, P. e Duarte Silva, P.), Edições SPE, pp. 271-284.

The Europeum (2012). Europe's University Budgets – The European Update, February 2012. *The Europeum – An Association of leading European Universities*. Disponível em: <http://www.europaeum.org/europaeum/?q=node/1636>.

## ANEXO 1. Glossário

Neste anexo apresentamos o glossário de algumas abreviações que foram usadas nesta dissertação.

Abreviação	Descrição
A	Artes
AI	Ano de Ingresso
AMU	Áreas Medianamente Urbanas
APR	Áreas Predominantemente Rurais
APU	Áreas Predominantemente Urbanas
ARL <sub>0</sub>	<i>Average Run Length</i>
AUC	<i>Area Under the Curve</i>
B	Biologia
C <sub>pku</sub>	Rácio da capacidade do processo cumprir especificações superiores, relativamente à média
CL	<i>Central Line</i>
CS	Ciências Sociais
CT	Ciências e Tecnologia
DC	Duração da parte Curricular
ECTS	<i>European Credit Transfer and accumulation System</i>
EMV	Estimativas de Máxima Verosimilhança
EWMA	<i>Exponentially Weighted Moving Average</i>
FC	Fase de Candidatura
GVPHC	Gestão e Valorização do Património Histórico e Cultural
HR	<i>Hazard Ratio</i> – razão de riscos
IC <sub>95%</sub>	Intervalos de confiança a 95%
KM	Kaplan-Meier
KS	Kolmogorov-Smirnov
LCL	<i>Lower Control Limit</i>
LSL	<i>Lower Specification Limit</i>
OR	<i>Odds Ratio</i> – razão de possibilidades
PCR	<i>Process Capability Ratio</i>
PE	Pedagogia da Educação
ROC	<i>Receiver Operating Characteristic Curve</i>
SAC	Serviços Académicos da Universidade de Évora
SN	<i>Skew-Normal</i>
SPC	<i>Statistical Process Control</i>
T diss	Tempo entre a data de aprovação do projeto e a data de entrega da dissertação
T formação júri	Tempo de formação do júri
T proc	Tempo entre a entrega da dissertação e a realização de provas públicas
T provas	Tempo desde a primeira reunião até à realização de provas públicas
T reunião	Tempo desde a formação do júri até à realização da primeira reunião
T sac	Tempo entre a entrega da dissertação e o seu envio para a escola
TD curso	Tempo de espera pela emissão de um diploma de curso
TD grau	Tempo de espera pela emissão de um diploma de grau
TE	Trabalhador Estudante
TPTI	Técnicas Patrimoniais e Territoriais da Indústria (História e Valorização)
TQM	<i>Total Quality Management</i>
TRV	Teste de Razão de Verosimilhanças
UCL	<i>Upper Control Limit</i>
USL	<i>Upper Specification Limit</i>

## ANEXO 2. Tabelas

**Tabela 1: Testes de normalidade KS e teste SN para o T proc ( $\leq 180$  dias) estratificado pelos anos de conclusão, escolas, departamentos e cursos. (\* distribuições para as quais assumimos uma distribuição SN)**

T proc ( $\leq 180$ dias)	Valor p (KS)	Valor p (SN)	Parâmetros da SN
2009 (Figura 5.14)	0,09	0,16	-
2010 (Figura 5.14)	0,02	0,95	-
2011 (Figura 5.14)	0,001	0,07	-
2012 (Figura 5.14)	0,18	0,49	-
2013 (Figura 5.14)	0,53	0,99	-
A (Figura 5.15)	0,55	0,85	-
CS (Figura 5.15)	0,001	0,06	-
CT (Figura 5.15)	0,12	0,94	-
Departamento E (Figura 5.16)	0,58	0,99	-
Departamento M (Figura 5.16)	0,24	0,89	-
Departamento PAO (Figura 5.17)	0,46	0,98	-
Departamento Q (Figura 5.17)	0,90	0,99	-
Departamento Z (Figura 5.17)	0,65	0,25	-
Departamento F (Figura 5.18)	0,04	<0,001*	$\lambda = 18,7; \delta = 55,9; \alpha = 722,2$
Curso P (Figura 5.19)	0,03	<0,001*	$\lambda = 175,7; \delta = 54,5; \alpha = -10,7$
Curso P ( $\leq 2011/2012$ ) (Figura 5.19)	0,002	<0,001*	$\lambda = 179,1; \delta = 43,8; \alpha = -861,2$
Curso P ( $> 2011/2012$ ) (Figura 5.19)	0,30	0,31	-
Departamento G (Figura 5.20)	0,02	0,99	-
Curso G (Figura 5.20)	0,02	0,99	-
Restantes cursos do departamento G (Figura 5.20)	0,04	0,23	-
Departamento H (Figura 5.21)	0,001	0,37	-
Curso GVPHC (Figura 5.21)	0,002	0,23	-
Curso TPTI (Figura 5.22)	0,001	0,06	-
Restantes cursos do departamento H (Figura 5.22)	0,13	0,82	-
Departamento PE (Figura 5.23)	0,32	0,99	-
Cursos de Ensino do departamento de PE (Figura 5.23)	0,31	0,78	-
Cursos de Ciências da Educação do departamento de PE (Figura 5.23)	0,33	0,99	-
Departamento S (Figura 5.24)	0,15	0,06	-
Departamento S ( $\leq 2010/2011$ ) (Figura 5.24)	0,04	0,01*	$\lambda = 170,2; \delta = 49,3; \alpha = -856,1$
Departamento S ( $> 2010/2011$ ) (Figura 5.24)	0,56	0,96	-
Departamento B (Figura 5.25)	0,79	0,99	-
Departamento B ( $\leq 2010/2011$ ) (Figura 5.25)	0,30	0,87	-
Departamento B ( $> 2010/2011$ ) (Figura 5.25)	0,41	0,25	-
Departamento DS (Figura 5.26)	0,48	0,07	-
Departamento DS ( $\leq 2011/2012$ ) (Figura 5.26)	0,25	0,95	-
Departamento DS ( $> 2011/2012$ ) (Figura 5.26)	0,15	0,06	-
Departamento I (Figura 5.27)	0,20	0,17	-
Departamento I ( $\leq 2010$ ) (Figura 5.27)	0,74	0,96	-
Departamento I ( $> 2010$ ) (Figura 5.27)	0,30	0,99	-

**Tabela 2: Testes de normalidade KS e teste SN para o T sac. (\* distribuições para as quais assumimos uma distribuição SN)**

	Valor p (KS)	Valor p (SN)	Parâmetros da SN
T sac (Figura 5.28)	<0,001	<0,001*	$\lambda = -0,08; \delta = 25,7; \alpha = 826,2$
$\leq$ Outubro de 2009 (Figura 5.29)	0,004	<0,001*	$\lambda = 0,9; \delta = 12,1; \alpha = 565,3$
> Outubro de 2009 (Figura 5.29)	<0,001	0,001*	$\lambda = 15,4; \delta = 12; \alpha = 2,8$
2010 (Figura 5.29)	<0,001	<0,001*	$\lambda = 28,2; \delta = 17,4; \alpha = 0,9$
> 2010 (Figura 5.30)	<0,001	<0,001*	$\lambda = 0,1; \delta = 23,2; \alpha = 676,4$
Escola A (Figura 5.31)	<0,001	0,02*	$\lambda = 0,88; \delta = 21,3; \alpha = 597,3$
Escola CS (Figura 5.31)	<0,001	<0,001*	$\lambda = 0,18; \delta = 19,4; \alpha = 47$

---

 ANEXOS
 

---

Escola CT (Figura 5.31)	<0,001	<0,001*	$\lambda = -0,11; \delta = 31,8; \alpha = 827,7$
T sac após 2010 e sem os departamentos DS, GVPC, I e B (Figura 5.32)	<0,001	<0,001*	$\lambda = 0,07; \delta = 17,7; \alpha = 691,4$

**Tabela 3: Testes de normalidade KS e teste SN para o T formação do júri. (\* distribuições para as quais assumimos uma distribuição SN)**

	Valor p (KS)	Valor p (SN)	Parâmetros da SN
T formação do júri (Figura 5.33)	<0,001	<0,001*	$\lambda = 7,7; \delta = 35,8; \alpha = 6$
Escola A (Figura 5.34)	0,20	0,95	-
Escola CS (Figura 5.34)	<0,001	<0,001*	$\lambda = 5,5; \delta = 41,4; \alpha = 16,3$
Escola CT (Figura 5.34)	<0,001	<0,001*	$\lambda = 7,8; \delta = 30,5; \alpha = 4,6$
2009 (Figura 5.35)	0,009	0,007*	$\lambda = 1,8; \delta = 30,5; \alpha = 13,1$
2010 (Figura 5.35)	<0,001	<0,001*	$\lambda = 13,1; \delta = 39,8; \alpha = 5,4$
> 2010 (Figura 5.35)	<0,001	<0,001*	$\lambda = 7,6; \delta = 34; \alpha = 5,8$
Finais de semestres (>2009) (Figura 5.36)	0,01	0,13	-
Restantes meses (>2009) (Figura 5.36)	<0,001	<0,001*	$\lambda = 10,9; \delta = 25; \alpha = 3,4$
Finais de semestres (2012 e 2013) (Figura 5.37)	0,02	0,06	-

**Tabela 4: Testes de normalidade KS e teste SN para o T reunião. (\* distribuições para as quais assumimos uma distribuição SN)**

	Valor p (KS)	Valor p (SN)	Parâmetros da SN
T reunião (Figura 5.38)	<0,001	<0,001*	$\lambda = -0,11; \delta = 32,9; \alpha = 882,8$
Escola A (Figura 5.39)	0,13	0,08	-
Escola CS (Figura 5.39)	<0,001	<0,001*	$\lambda = 1,12; \delta = 35,9; \alpha = 12,6$
Escola CT (Figura 5.39)	<0,001	<0,001*	$\lambda = -0,10; \delta = 27; \alpha = 822,7$
2009 (Figura 5.40)	0,007	0,001*	$\lambda = -0,06; \delta = 11,9; \alpha = 666$
2010 (Figura 5.40)	<0,001	<0,001*	$\lambda = -0,13; \delta = 34,2; \alpha = 813,7$
2011 (Figura 5.40)	<0,001	<0,001*	$\lambda = 0,04; \delta = 31,9; \alpha = 81$
2012 (Figura 5.41)	<0,001	<0,001*	$\lambda = -0,11; \delta = 33,7; \alpha = 876,1$
2013 (Figura 5.41)	<0,001	<0,001*	$\lambda = 0,9; \delta = 34,8; \alpha = 967,4$
T reunião inferior a 100 dias (Figura 5.42)	<0,001	<0,001*	$\lambda = -0,11; \delta = 32,3; \alpha = 887,2$

**Tabela 5: Testes de normalidade KS e teste SN para o T provas. (\* distribuições para as quais assumimos uma distribuição SN)**

	Valor p (KS)	Valor p (SN)	Parâmetros da SN
T provas (Figura 5.43)	<0,001	<0,001*	$\lambda = 2,8; \delta = 30,6; \alpha = 12,5$
Escola A (Figura 5.44)	0,13	0,08	-
Escola CS (Figura 5.44)	<0,001	<0,001*	$\lambda = 4; \delta = 32,1; \alpha = 10,1$
Escola CT (Figura 5.44)	<0,001	<0,001*	$\lambda = 2; \delta = 27,4; \alpha = 14,6$
2009 (Figura 5.45)	0,003	0,02*	$\lambda = 1,2; \delta = 117,8; \alpha = 5,7$
2010 (Figura 5.45)	<0,001	<0,001*	$\lambda = 0,5; \delta = 22,6; \alpha = 27,8$
2011 (Figura 5.45)	<0,001	<0,001*	$\lambda = 5,4; \delta = 26,3; \alpha = 6,6$
2012 (Figura 5.46)	<0,001	<0,001*	$\lambda = 4,4; \delta = 33,3; \alpha = 14,1$
2013 (Figura 5.46)	<0,001	<0,001*	$\lambda = 5,9; \delta = 32,6; \alpha = 13,6$

**Tabela 6: Testes de normalidade KS e teste SN para o TD grau e TD curso. (\* distribuições para as quais assumimos uma distribuição SN)**

	Valor p (KS)	Valor p (SN)	Parâmetros da SN
TD grau (Figura 5.47)	<0,001	<0,001*	$\lambda = -0,07; \delta = 9,9; \alpha = 183,5$
TD grau 2010 (Figura 5.48)	<0,001	<0,001*	$\lambda = 0,9; \delta = 12,8; \alpha = 19,6$
TD grau 2011 (Figura 5.48)	<0,001	<0,001*	$\lambda = -0,07; \delta = 10,5; \alpha = 183,5$

---

---

## ANEXOS

---

TD grau 2012 (Figura 5.49)	<0,001	<0,001*	$\lambda = -0,06; \delta = 9; \alpha = 183,5$
TD grau 2013 (Figura 5.49)	<0,001	<0,001*	$\lambda = -0,08; \delta = 8,1; \alpha = 183,5$
TD curso (Figura 5.50)	<0,001	<0,001*	$\lambda = -0,08; \delta = 9,1; \alpha = 183,5$
TD curso 2010 (Figura 5.51)	0,07	0,20	-
TD curso 2011 (Figura 5.52)	<0,001	<0,001*	$\lambda = -0,09; \delta = 9,6; \alpha = 183,5$
TD curso 2012 (Figura 5.52)	<0,001	<0,001*	$\lambda = -0,04; \delta = 8,3; \alpha = 183,5$
TD curso 2013 (Figura 5.52)	<0,001	<0,001*	$\lambda = -0,08; \delta = 9,1; \alpha = 183,5$

---