

**UNIVERSIDADE ABERTA**



## **Análise de Sobrevivência**

**Uma Aplicação a Eventos de Golo em Jogos de Futebol**

Joaquim Manuel Jorge Borges

Mestrado em Estatística, Matemática e Computação

2018

**UNIVERSIDADE ABERTA**



## **Análise de Sobrevivência**

**Uma Aplicação a Eventos de Golo em Jogos de Futebol**

Joaquim Manuel Jorge Borges

Mestrado em Estatística, Matemática e Computação

Dissertação orientada pelo Prof. Dr. Pedro Serranho

2018

# Resumo

Este estudo tem como principal objetivo caracterizar e comparar a sobrevivência a eventos de golo em jogos de futebol estendendo os resultados de trabalhos anteriormente publicados (*Nevo e Ritov, 2012*) a um conjunto mais alargado de ligas e a diferentes eventos de golo.

Centra-se, assim, na aplicação de técnicas estatísticas da área da Análise de Sobrevivência, subárea da Estatística, que se dedica à definição de métodos e modelos destinados à análise de dados de sobrevivência, isto é, dados em que a principal variável em estudo é o tempo até ao evento.

A análise baseou-se em dados relativos a 5 épocas (2012-2017) de 6 ligas europeias (Portugal, Espanha, Alemanha, Inglaterra, França, Itália).

Para estes dados foram calculadas as estimativas de Kaplan-Meier da função de sobrevivência e foi feito o ajustamento dos dados ao modelo de riscos proporcionais de Cox.

Os eventos de interesse considerados foram os eventos de marcador correspondentes ao 1.º golo, 2.º golo e 3.º golo.

Concluimos que, relativamente ao 1.º golo, à semelhança dos resultados de outros autores, o fator casa é de facto determinante para a definição do resultado final.

Por outro lado, a taxa de falha do 2.º golo, quando marcado pela equipa de casa estando esta a vencer, revelou que o 1.º golo induz a ocorrência de 2.º golo mas este ocorre com maior probabilidade após um intervalo de 45-50 minutos.

No caso do 3.º golo, concentrámo-nos nos eventos de golo correspondentes ao desempate, onde constatámos uma tendência de crescimento ao longo do tempo muito semelhante para os dois tipos de eventos: golo da equipa de casa (2-1), golo da equipa de fora (1-2).

Alguns dos aspetos que poderão ser objeto de estudos futuros incluem o enriquecimento com inclusão de mais covariáveis, a extensão a um maior número de ligas e épocas desportivas e o recurso à extensão do modelo de Cox com uma variável fragilidade.

**Palavras-chave:** Jogo de Futebol, Golo, Análise de Sobrevivência, Taxa de Falha, Modelo de Cox

# Abstract

The aim of this study is to portray and compare survival functions of scoring events in soccer matches by extending the results of previously published works (Nevo and Ritov, 2012) to a wider set of leagues and to more scoring events.

Our focus was on the application of techniques from Survival Analysis, a sub-area of Statistics, which is dedicated to the definition of methods of analysis and modelling of survival data, i.e., data in which the main variable under study is the time to the event.

The analysis was based on data from 5 seasons (2012-2017) of 6 European leagues (Portugal, Spain, Germany, England, France, Italy).

For these data we obtained the Kaplan-Meier estimates of the survival function and we adjusted the Cox proportional hazards model.

The events of interest corresponded to the scoring events in reference to the 1st goal, the 2nd goal and the 3rd goal.

We concluded that, for the 1st goal, as concluded by other authors, the home team factor is in fact decisive for the making of the final score.

On the other hand, the failure rate of the 2nd goal, scored by the winning home team, revealed that the 1st goal induces a second goal, but this occurs more likely after an interval of 45-50 minutes.

In the case of the third goal, we focused on tiebreaking events, and we observed a trend of growth over time very similar for the two types of events: home team goal (2-1), visiting team goal (1-2).

Some of the aspects that may be a subject for future studies include enrichment with inclusion of more covariables, extension to a greater number of sports leagues and seasons and the use of an extension of the Cox model with a frailty variable.

**Keywords:** Football Match, Goal, Survival Analysis, Failure Rate, Cox Model

# Dedicatória

à minha companheira, Guiomar Fragoso

# Agradecimentos

Agradeço ao Prof. Pedro Serranho pela paciência e entusiasmo com que acompanhou e orientou esta dissertação.

# Índice

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Revisão sobre Análise de Sobrevivência</b>	<b>3</b>
2.1	Conceitos de Análise de Sobrevivência . . . . .	4
2.1.1	Censura e Truncatura . . . . .	4
2.1.2	Variável Principal e Evento de Interesse . . . . .	5
2.1.3	Covariáveis . . . . .	5
2.2	Estimadores Não Paramétricos da Função de Sobrevivência . . . . .	6
2.2.1	Comparação de Funções de Sobrevivência . . . . .	9
2.3	Modelo de Cox . . . . .	10
2.3.1	Componentes do Modelo de Cox . . . . .	10
2.3.2	Testes da Hipótese de Riscos Proporcionais . . . . .	12
2.3.3	Seleção de Variáveis do Modelo . . . . .	13
2.4	Aplicações da Análise de Sobrevivência no Âmbito de Fenómenos Desportivos . . . . .	14
<b>3</b>	<b>Metodologia</b>	<b>17</b>
3.1	Metodologia de Investigação . . . . .	17
3.2	Metodologia de Recolha de Dados . . . . .	17
3.3	Armazenamento de Dados . . . . .	18
3.4	Software Estatístico . . . . .	18
3.5	Análise Estatística . . . . .	18
3.5.1	Eventos de Interesse . . . . .	19
3.5.2	Estimador da Função de Sobrevivência - Kaplan-Meier . . . . .	19
3.5.3	Modelo de Cox . . . . .	20
<b>4</b>	<b>Apresentação de Resultados</b>	<b>21</b>
4.1	1.º Golo da Equipa . . . . .	21
4.1.1	Estimativa de Kaplan-Meier da Função de Sobrevivência de Equipa de Casa Marca Respetivo 1.º Golo . . . . .	22
4.1.2	Estimativa de Kaplan-Meier da Função de Sobrevivência de Equipa de Fora Marca Respetivo 1.º Golo . . . . .	22
4.2	1.º Golo do Jogo . . . . .	23
4.2.1	Estimativa de Kaplan-Meier da Função de Sobrevivência de 1.º Golo Marcado pela Equipa de Casa . . . . .	23
4.2.2	Estimativa de Kaplan-Meier da Função de Sobrevivência de 1.º Golo Marcado pela Equipa de Fora . . . . .	26

4.3	Modelo de Cox para o 1.º Golo do Jogo . . . . .	27
4.3.1	Hipótese de Riscos Proporcionais . . . . .	28
4.3.2	Equipa de Casa . . . . .	29
4.3.3	Equipa de Fora . . . . .	31
4.3.4	Comparação das Funções Taxa de Falha: Equipa de Casa vs. Equipa de Fora . . . . .	35
4.4	Estimativa de Kaplan-Meier para o 2.º Golo . . . . .	35
4.4.1	2.º Golo do Jogo . . . . .	35
4.4.2	Equipa de Casa . . . . .	36
4.4.3	Equipa de Fora . . . . .	37
4.5	Modelo de Cox para o 2.º Golo . . . . .	38
4.5.1	Hipótese de Riscos Proporcionais . . . . .	38
4.5.2	Equipa de Casa: de 1-0 para 2-0 . . . . .	39
4.5.3	Equipa de Casa: de 0-1 para 1-1 . . . . .	40
4.5.4	Equipa de Fora: de 1-0 para 1-1 . . . . .	42
4.5.5	Equipa de Fora: de 0-1 para 0-2 . . . . .	44
4.5.6	Comparação de Taxas de Falha: 1-0 para 2-0 vs. 1-0 para 1-1 . . . . .	47
4.5.7	Comparação de Taxas de Falha: 0-1 para 0-2 vs. 0-1 para 1-1 . . . . .	47
4.6	Estimativa de Kaplan-Meier da Função de Sobrevivência de 2.º Golo do Jogo com Tempo Inicial Igual ao Tempo do 1.º Golo . . . . .	48
4.6.1	2.º Golo do Jogo com Tempo Inicial Igual ao Tempo do 1.º Golo . . . . .	49
4.6.2	2.º Golo do Jogo Marcado por Equipa de Casa com Tempo Inicial Igual ao Tempo do 1.º Golo . . . . .	50
4.6.3	2.º Golo do Jogo Marcado por Equipa de Fora com Tempo Inicial Igual ao Tempo do 1.º Golo . . . . .	51
4.7	Modelo de Cox para o 2.º Golo do Jogo com Tempo Inicial Igual a Tempo do 1.º Golo . . . . .	52
4.7.1	Hipótese de Riscos Proporcionais . . . . .	52
4.7.2	Equipa de Casa: de 1-0 para 2-0 . . . . .	53
4.7.3	Equipa de Casa: de 0-1 para 1-1 . . . . .	54
4.7.4	Equipa de Casa: de 1-0 para 1-1 . . . . .	56
4.7.5	Equipa de Fora: de 0-1 para 0-2 . . . . .	58
4.7.6	Comparação de Funções Taxa de Falha: 1-0 para 2-0 vs. 1-0 para 1-1 . . . . .	61
4.7.7	Comparação de Funções Taxa de Falha: 0-1 para 0-2 vs. 0-1 para 1-1 . . . . .	61
4.8	Estimativa de Kaplan-Meier para o 3.º Golo do Jogo . . . . .	62
4.8.1	3.º Golo do Jogo . . . . .	62
4.8.2	3.º Golo Marcado pela Equipa de Casa . . . . .	63
4.8.3	3.º Golo do Jogo Marcado por Equipa de Fora . . . . .	65
4.9	Modelo de Cox para o 3.º Golo do Jogo . . . . .	65
4.9.1	Hipótese de Riscos Proporcionais . . . . .	65
4.9.2	Equipa de Casa: de 1-1 para 2-1 . . . . .	67
4.9.3	Equipa de Fora: de 1-1 para 1-2 . . . . .	69
4.9.4	Comparação das Funções Taxa de Falha de 1-1 para 2-1 vs. de 1-1 para 1-2 . . . . .	70



4.10	Estimativa de Kaplan-Meier do 3.º Golo do Jogo com Tempo Inicial Igual ao Tempo do 2.º Golo . . . . .	70
4.10.1	3.º Golo do Jogo . . . . .	72
4.10.2	3.º Golo Marcado por Equipa de Casa . . . . .	73
4.10.3	3.º Golo Marcado por Equipa de Fora . . . . .	74
4.11	Modelo de Cox para o 3.º Golo do Jogo com Tempo Inicial Igual a Tempo do 2.º Golo . . . . .	75
4.11.1	Hipótese de Riscos Proporcionais . . . . .	75
4.11.2	Equipa de Casa: de 1-1 para 2-1 . . . . .	76
4.11.3	Equipa de Fora: de 1-1 para 1-2 com $t_0 = t_{golo2}$ . . . . .	77
4.11.4	Comparação de Função Taxa de Falha de 1-2 para 2-1 vs. 1-1 para 1-2 com $t_0 = t_{golo2}$ . . . . .	79
<b>5</b>	<b>Conclusões</b>	<b>81</b>
5.1	1.º Golo . . . . .	81
5.2	2.º Golo . . . . .	82
5.3	3.º Golo . . . . .	83
5.4	Perspetivas de Trabalho de Futuro . . . . .	83

# Lista de Figuras

4.1	Estimativa de Kaplan-Meier da Função de Sobrevivência de Equipa de Casa Marca o Respetivo 1.º Golo . . . . .	23
4.2	Estimativa de Kaplan-Meier da Função de Sobrevivência de Equipa de Fora Marca o Respetivo 1.º Golo . . . . .	24
4.3	Estimativa de Kaplan-Meier da Função de Sobrevivência de 1.º Golo Marcado por Equipa de Casa . . . . .	25
4.4	Estimativa de Kaplan-Meier de 1.º Golo Marcado por Equipa de Fora . . . . .	27
4.5	Resíduos de Schoenfeld para Modelo de Cox do 1.º Golo do Jogo . . . . .	28
4.6	Função de Sobrevivência pelo Modelo de Cox para 1.ºGolo-Casa: Covariável ligas, odds fixa. . . . .	30
4.7	Função de Sobrevivência pelo Modelo de Cox para 1.ºGolo-Casa: Covariável odds, liga fixa . . . . .	30
4.8	Função Taxa de Falha basal para 1.ºGolo- Equipa de Casa . . . . .	31
4.9	Função Taxa de Falha Acumulada para 1.ºGolo-Equipa de Casa . . . . .	32
4.10	Função de Sobrevivência pelo Modelo de Cox para 1.ºGolo - Fora: Covariável ligas, odds fixa . . . . .	33
4.11	Função de Sobrevivência pelo Modelo de Cox para 1.ºGolo - Fora: Covariável odds, liga fixa . . . . .	33
4.12	Função Taxa de Falha basal para 1.ºGolo - Equipa de Fora . . . . .	34
4.13	Função Taxa de Falha Acumulada para 1.ºGolo - Fora . . . . .	34
4.14	Comparação de Funções Taxa de Falha basal para 1.ºGolo . . . . .	35
4.15	Estimativa de Kaplan-Meier da Função de Sobrevivência para 2.º Golo do Jogo . . . . .	36
4.16	Estimativa de Kaplan-Meier da Função de Sobrevivência para 2.º Golo do Jogo Marcado pela Equipa de Casa . . . . .	37
4.17	Estimativa de Kaplan-Meier da Função de Sobrevivência para 2.º Golo do Jogo Marcado pela Equipa de Fora . . . . .	38
4.18	Resíduos de Schoenfeld para Modelo de Cox do 2.º Golo do Jogo . . . . .	39
4.19	Taxa de Falha do 2.º Golo do Jogo: de 1-0 para 2-0 . . . . .	41
4.20	Taxa de Falha Acumulada do 2.º Golo do Jogo: de 1-0 para 2-0 . . . . .	41
4.21	Função Taxa de Falha de 2.º Golo do Jogo: de 0-1 para 1-1 . . . . .	43
4.22	Função Taxa de Falha de 2.º Golo do Jogo: de 0-1 para 1-1 . . . . .	43
4.23	Função Taxa de Falha de 2.º Golo do Jogo: de 1-0 para 1-1 . . . . .	45
4.24	Função Taxa de Falha Acumulada de 2.º Golo do Jogo: de 1-0 para 1-1 . . . . .	45
4.25	Função Taxa de Falha de 2.º Golo do Jogo: de 0-1 para 0-2 . . . . .	46
4.26	Função Taxa de Falha Acumulada de 2.º Golo do Jogo: de 0-1 para 0-2 . . . . .	47
4.27	Comparação de Taxas de Falha: 1-0 para 2-0 vs. 1-0 para 1-1 . . . . .	48

4.28	Comparação de Taxas de Falha: 0-1 para 0-2 vs. 0-1 para 1-1 . . . . .	48
4.29	Estimativa de Kaplan-Meier da Função de Sobrevivência de 2.º Golo do Jogo com $t_0=t_{golo1}$ . . . . .	50
4.30	Estimativa de Kaplan-Meier da Função de Sobrevivência de 2.º Golo do Jogo com $t_0=t_{golo1}$ . . . . .	51
4.31	Estimativa de Kaplan-Meier da Função de Sobrevivência de 2.º Golo do Jogo com $t_0=t_{golo1}$ . . . . .	52
4.32	Resíduos de Schoenfeld para 2.º golo com $t_0 = t_{golo1}$ . . . . .	53
4.33	Função Taxa de Falha para 2.º Golo de Jogo com $t_0 = t_{golo1}$ . . . . .	55
4.34	Função Taxa de Falha Acumulada para 2.º Golo de Jogo com $t_0 = t_{golo1}$	55
4.35	Função Taxa de Falha de 2.º Golo: de 0-1 para 1-1 com $t_0 = t_{golo2}$ . . .	56
4.36	Função Taxa de Falha de 2.º Golo: de 0-1 para 1-1 com $t_0 = t_{golo2}$ . . .	57
4.37	Função Taxa de Falha de 2.º Golo: de 1-0 para 1-1 com $t_0 = t_{golo2}$ . . .	58
4.38	Função Taxa de Falha de 2.º Golo: de 1-0 para 1-1 com $t_0 = t_{golo2}$ . . .	59
4.39	Função Taxa de Falha de 2.º Golo: de 0-1 para 0-2 com $t_0 = t_{golo2}$ . . .	60
4.40	Função Taxa de Falha de 2.º Golo: de 0-1 para 0-2 com $t_0 = t_{golo2}$ . . .	60
4.41	Comparação de Funções Taxa de Falha de 2.º Golo: de 1-0 para 2-0 vs. 1-0 para 1-1 com $t_0 = t_{golo1}$ . . . . .	61
4.42	Comparação de Funções Taxa de Falha de 2.º Golo: de 0-1 para 0-2 vs. 0-1 para 1-1 com $t_0 = t_{golo2}$ . . . . .	62
4.43	Estimativa de Kaplan-Meier da Função de Sobrevivência de 3.º Golo do Jogo . . . . .	63
4.44	Estimativa de Kaplan-Meier da Função de Sobrevivência de 3.º Golo do Jogo Marcado por Equipa de Casa . . . . .	64
4.45	Estimativa de Kaplan-Meier da Função de Sobrevivência 3.º Golo Marcado por Equipa de Fora . . . . .	66
4.46	Resíduos de Schoenfeld de 3.º Golo do Jogo . . . . .	66
4.47	Função Taxa de Falha de 3.º Golo: de 1-1 para 2-1 . . . . .	68
4.48	Função Taxa de Falha Acumulada de 3.º Golo: de 1-1 para 2-1 . . . . .	69
4.49	Função Taxa de Falha de 3.º Golo: de 1-1 para 1-2 . . . . .	71
4.50	Função Taxa de Falha Acumulada de 3.º Golo: de 1-1 para 1-2 . . . . .	71
4.51	Comparação das Funções Taxa de Falha de 1-1 para 2-1 vs. de 1-1 para 1-2 . . . . .	72
4.52	Estimativa de Kaplan-Meier da Função de Sobrevivência 3.º Golo do Jogo com $t_0 = t_{golo2}$ . . . . .	73
4.53	Estimativa de Kaplan-Meier da Função de Sobrevivência 3.º Golo do Jogo Marcado por Equipa de Casa com $t_0 = t_{golo2}$ . . . . .	74
4.54	Estimativa de Kaplan-Meier da Função de Sobrevivência 3.º Golo do Jogo Marcado por Equipa de Fora com $t_0 = t_{golo2}$ . . . . .	75
4.55	Resíduos de Schoenfeld para 3.º Golo do Jogo com $t_0 = t_{golo2}$ . . . . .	76
4.56	Função Taxa de Falha de 3.º Golo: de 1-1 para 2-1 com $t_0 = t_{golo2}$ . . .	77
4.57	Função Taxa de Falha de 3.º Golo: de 1-1 para 2-1 com $t_0 = t_{golo2}$ . . .	78
4.58	Função Taxa de Falha de 3.º Golo: de 1-1 para 1-2 com $t_0 = t_{golo2}$ . . .	79
4.59	Função Taxa de Falha de 3.º Golo: de 1-1 para 1-2 com $t_0 = t_{golo2}$ . . .	79
4.60	Compara de Funções Taxa de Falha de 3.º Golo: 1-1 para 1-2 vs. 1-2 para 2-1 com $t_0 = t_{golo2}$ . . . . .	80

# Lista de Tabelas

2.1	Tabela de Frequências para o Teste de Log-Rank . . . . .	9
4.1	Quartis de Tempo de Sobrevivência para Equipa de Casa Marca Respe- tivo 1.º Golo . . . . .	22
4.2	Quartis de Tempo de Sobrevivência para Equipa de Fora Marca Respe- tivo 1.º Golo . . . . .	23
4.3	Quartis de Tempo de Sobrevivência para a Estimativa de Kaplan-Meier de 1.º Golo Marcado pela Equipa de Casa . . . . .	25
4.4	Quartis de Tempo de Sobrevivência para a Estimativa de Kaplan-Meier de 1.º Golo Marcado pela Equipa de Fora . . . . .	26
4.5	Listagem de Variáveis . . . . .	27
4.6	Modelo de Cox para 1.º Golo - Casa: modelo inicial . . . . .	29
4.7	Modelo de Cox para 1.º Golo - Casa: Passos de Eliminação de Covariáveis	29
4.8	Modelo de Cox para 1.º Golo - Casa: modelo final . . . . .	29
4.9	Modelo de Cox para 1.º Golo - Fora: modelo inicial . . . . .	32
4.10	Modelo de Cox para 1.º Golo - Fora: Passos de Seleção de Covariáveis .	32
4.11	Modelo de Cox para 1.º Golo - Fora: modelo final . . . . .	32
4.12	Quartis do Tempo de Sobrevivência do 2.º Golo do Jogo . . . . .	36
4.13	Quartis do Tempo de Sobrevivência do 2.º Golo do Jogo Marcado pela Equipa de Casa . . . . .	37
4.14	Quartis do Tempo de Sobrevivência do 2.º Golo do Jogo Marcado pela Equipa de Fora . . . . .	37
4.15	Modelo de Cox para 2.º Golo do Jogo: de 1-0 para 2-0 - modelo inicial	40
4.16	Modelo de Cox para 2.º Golo: de 1-0 para 2-0 - Passos de Seleção de Covariáveis . . . . .	40
4.17	Modelo de Cox para 2.º Golo do Jogo: de 1-0 para 2-0 - modelo final .	40
4.18	Modelo de Cox para 2.º Golo do Jogo: de 1-0 para 2-0 - modelo inicial	42
4.19	Modelo de Cox para 2.º Golo: de 0-1 para 1-1 - Passos de Seleção de Covariáveis . . . . .	42
4.20	Modelo de Cox para 2.º Golo do Jogo: de 0-1 para 1-1 - modelo final .	42
4.21	Modelo de Cox para 2.º Golo do Jogo: de 1-0 para 1-1 - modelo inicial	44
4.22	Modelo de Cox para 2.º Golo: de 1-0 para 1-1 - Passos de Seleção de Covariáveis . . . . .	44
4.23	Modelo de Cox para 2.º Golo do Jogo: de 1-0 para 1-1 - modelo final .	44
4.24	Modelo de Cox para 2.º Golo do Jogo: de 0-1 para 0-2 - modelo inicial	46
4.25	Modelo de Cox para 2.º Golo: de 0-1 para 0-2 - Seleção de Covariáveis	46
4.26	Quartis de Tempo de Sobrevivência de 2.º Golo do Jogo com $t_0 = t_{golo1}$	49

4.27	Quartis de Tempo de Sobrevivência de 2.º Golo do Jogo Marcado por Equipa de Casa com $t_0 = t_{golo1}$ . . . . .	50
4.28	Quartis de Tempo de Sobrevivência de 2.º Golo do Jogo Marcado por Equipa de Fora com $t_0 = t_{golo1}$ . . . . .	51
4.29	Modelo de Cox para 2.º Golo do Jogo de 1-0 para 2-0 com $t_0 = t_{golo1}$ : modelo inicial . . . . .	53
4.30	Modelo de Cox para 2.º Golo do Jogo de 1-0 para 2-0 com $t_0 = t_{golo1}$ : Passos de Seleção de Covariáveis . . . . .	54
4.31	Modelo de Cox para 2.º Golo do Jogo de 1-0 para 2-0 com $t_0 = t_{golo1}$ : modelo final . . . . .	54
4.32	Modelo de Cox para 2.º Golo do Jogo de 0-1 para 1-1 com $t_0 = t_{golo1}$ : modelo inicial . . . . .	56
4.33	Modelo de Cox para 2.º Golo do Jogo de 0-1 para 1-1 com $t_0 = t_{golo1}$ : Passos de Seleção de Covariáveis . . . . .	56
4.34	Modelo de Cox para 2.º Golo do Jogo de 1-0 para 1-1 com $t_0 = t_{golo1}$ : modelo inicial . . . . .	57
4.35	Modelo de Cox para 2.º Golo do Jogo de 0-1 para 1-1 com $t_0 = t_{golo1}$ : Passos de Seleção de Covariáveis . . . . .	57
4.36	Modelo de Cox para 2.º Golo do Jogo de 1-0 para 1-1 com $t_0 = t_{golo1}$ : modelo final . . . . .	58
4.37	Modelo de Cox para 2.º Golo do Jogo de 0-1 para 0-2 com $t_0 = t_{golo1}$ : modelo inicial . . . . .	59
4.38	Modelo de Cox para 2.º Golo do Jogo de 0-1 para 0-2 com $t_0 = t_{golo1}$ : Passos de Seleção de Covariáveis . . . . .	59
4.39	Quartis de Tempo de Sobrevivência para 3.º Golo do Jogo . . . . .	63
4.40	Quartis de Tempo de Sobrevivência para 3.º Golo do Jogo Marcado por Equipa de Casa . . . . .	64
4.41	Quartis de Tempo de Sobrevivência para 3.º Golo do Jogo Marcado por Equipa de Fora . . . . .	65
4.42	Modelo de Cox para 3.º Golo do Jogo: de 1-1 para 2-1 - modelo inicial	67
4.43	Modelo de Cox para 3.º Golo do Jogo: de 1-1 para 2-1 - Passos de Seleção de Covariáveis . . . . .	67
4.44	Modelo de Cox para 3.º Golo do Jogo: de 1-1 para 2-1 - modelo final . . . . .	68
4.45	Modelo de Cox para 3.º Golo do Jogo: de 1-1 para 1-2 - modelo inicial	70
4.46	Modelo de Cox para 3.º Golo do Jogo: de 1-1 para 1-2 - Passos de Seleção de Covariáveis . . . . .	70
4.47	Modelo de Cox para 3.º Golo do Jogo: de 1-1 para 1-2 - modelo final . . . . .	70
4.48	Quartis de Tempo de Sobrevivência para 3.º Golo do Jogo com $t_0 = t_{golo2}$	72
4.49	Quartis de Tempo de Sobrevivência para 3.º Golo do Jogo Marcado por Equipa de Casa com $t_0 = t_{golo2}$ . . . . .	73
4.50	Quartis de Tempo de Sobrevivência para 3.º Golo do Jogo Marcado por Equipa de Casa com $t_0 = t_{golo2}$ . . . . .	74
4.51	Modelo de Cox para 3.º Golo do Jogo: de 1-1 para 2-1 com $t_0 = t_{golo2}$ - modelo inicial . . . . .	76
4.52	Modelo de Cox para 3.º Golo do Jogo: de 1-1 para 2-1 com $t_0 = t_{golo2}$ - Passos de Seleção de Covariáveis . . . . .	77

4.53	Modelo de Cox para 3.º Golo do Jogo: de 1-1 para 1-2 com $t_0 = t_{golo2}$ - modelo inicial . . . . .	78
4.54	Modelo de Cox para 3.º Golo do Jogo: de 1-1 para 1-2 com $t_0 = t_{golo2}$ - Passos de Seleção de Covariáveis . . . . .	78

# Capítulo 1

## Introdução

O estudo que aqui se apresenta tem como principal objetivo caracterizar e comparar a sobrevivência a eventos de golo em jogos de futebol, estendendo trabalhos anteriormente publicados a um conjunto mais alargado de ligas e a diferentes eventos de golo.

Dessa forma, pretendeu-se estimar a função de sobrevivência e identificar e analisar a forma como diversas variáveis influenciam o comportamento da função de sobrevivência de eventos de golo de jogos de futebol.

Este estudo tem o seu foco na aplicação de técnicas da Análise de Sobrevivência, subárea da Estatística que se dedica à definição de "métodos e modelos destinados" à análise de dados de sobrevivência (*Rocha e Papoila, 2008*).

Na concretização deste estudo entendemos que o uso de técnicas de Análise de Sobrevivência trará uma nova luz para o entendimento dos fenómenos no âmbito das atividades desportivas e dessa forma alargará o âmbito de aplicação das técnicas estatísticas, em particular da Análise de Sobrevivência.

Assim este estudo assemelha-se, na sua intenção, a outros estudos que podem ser entendidos como percursos do uso de um conjunto de técnicas estatísticas que permitam uma melhor preparação dos técnicos e treinadores de futebol e uma melhoria na performance das suas equipas.

Por outro lado, a realização deste estudo será um contributo modesto para permitir à comunidade das apostas desportivas dispor de ferramentas que caracterizem a sua atividade principal. Nomeadamente concretizando-se numa aplicação informática que, através de uma atualização constante, dê conta dos resultados aqui obtidos e de outros resultados provenientes de outras ligas e outras épocas desportivas.

No domínio da análise de dados de sobrevivência define-se como principal variável de estudo o tempo até ao evento, neste caso o golo, e a característica mais distintiva destes dados é o facto de em geral conterem tempos de vida censurados, ou seja identificados com um valor de que apenas se sabe que é inferior ao tempo de vida efetivo.

As principais técnicas aplicadas em Análise de Sobrevivência consistem na estimação da função de sobrevivência e na definição de um modelo de regressão (denominado modelo de Cox) que identifique as principais variáveis que explicam a probabilidade de ocorrência do evento de interesse.

Para a análise dos dados relativos a eventos de golo, realizada no presente estudo, foram consideradas 6 ligas europeias (Portugal, Espanha, Alemanha, Inglaterra, França,

Itália) e considerados os jogos das cinco épocas desportivas mais recentes (2012-2013 a 2016-2017) todos com dados disponíveis em documentos online.

A ocorrência de eventos de golo num jogo de futebol pode significar um número elevado pelo que podem desenvolver-se várias combinações na marcha do marcador para o resultado final, principal motivo porque nos limitámos a analisar os 3 primeiros eventos de golo (1.º golo, 2.º golo e 3.º golo).

No caso do 3.º golo considerámos apenas os eventos de golo correspondentes ao desempate no marcador (de 1-1 para 2-1 e de 1-1 para 1-2), por nos parecerem os mais relevantes.

Nos diversos eventos foi possível identificar as variáveis que respeitavam os pressupostos para a construção do modelo de Cox (modelo de riscos proporcionais) e calcular os valores dos coeficientes associados às diversas variáveis, assim como estimar a função de sobrevivência usando métodos não paramétricos como o estimador de Kaplan-Meier.

Na sequência desta apresentação do tema passamos a descrever a organização da dissertação. Assim, o capítulo 2 é dedicado à revisão sobre as técnicas de Análise de Sobrevivência, incluindo uma revisão da literatura dedicada às aplicações da Análise de Sobrevivência no âmbito do Desporto.

No capítulo 3 apresentamos a metodologia do estudo indicando os procedimentos de recolha e armazenamento de dados e explicitamos também a metodologia estatística aplicada.

No capítulo 4 são apresentados os resultados obtidos quanto às estimativas e modelos de regressão ajustados aos dados dos 3 tipos de eventos (1.º golo, 2.º golo e 3.º golo), bem como as respetivas funções taxa de falha.

No capítulo 5 apresentaremos as principais conclusões para cada um dos eventos e a comparação de alguns resultados, obtidos relativamente a tipos de eventos do mesmo tipo, mas que representam uma configuração do marcador diferente.

Neste último capítulo adiantaremos também alguns aspetos a considerar em abordagens futuras na aplicação da Análise de Sobrevivência a eventos de golo de jogos de futebol.



## Capítulo 2

# Revisão sobre Análise de Sobrevivência

Neste capítulo apresentamos uma breve análise dos principais conceitos e resultados que se podem encontrar em diversa literatura dedicada à Análise de Sobrevivência em particular, terminaremos depois com uma enumeração de algumas aplicações a eventos desportivos.

A Análise de Sobrevivência é a área da Estatística dedicada à análise de dados referentes ao "tempo de vida", usando por esse motivo técnicas que procuram analisar e resolver problemas relacionados com dados em que o "tempo – at – ao – evento" é o elemento central.

Tratam-se de dados que podem ser complexos, não só pela natureza dinâmica dos eventos em estudo, mas também porque, muito frequentemente, incluem dados censurados nomeadamente quando a ocorrência do evento de interesse não é observada diretamente (*Rocha e Papoila, 2008*).

Historicamente esta disciplina estava associada às Ciências Atuariais e à Demografia (*Rocha e Papoila, 2008*), com origem nos séc. XVII e XVIII. Mais recentemente conheceu dois momentos fundamentais de desenvolvimento significativo que se têm revelado decisivos no estabelecimento de padrões de investigação de problemas de diversos estudos na área da Saúde.

Surgindo como associado ao estudo da sobrevivência dos indivíduos sujeitos a tratamentos do cancro, um desses momentos consistiu na proposta, por Kaplan e Meier em 1958 (*Kaplan e Meier, 1958*), de um estimador não paramétrico da função de sobrevivência - função que faz corresponder um determinado tempo de vida a uma probabilidade de sobrevivência.

Um outro momento, em 1972, compreendeu a definição de um modelo de regressão de cariz semi-paramétrico para a função de risco - função que faz corresponder um determinado tempo de vida com uma taxa ou variação instantânea de morte (ou falha).

O seu autor, David Cox, teve como objetivo estender os resultados obtidos para o estimador de Kaplan-Meier à comparação de tabelas de vida e incorporar argumentos semelhantes aos usados nos modelos de regressão aplicando-os à análise do tempo de vida (*Cox, 1972*).

De forma abreviada podemos dizer que esta área da Estatística se tem dedicado a desenvolver métodos para a análise de dados que, de forma essencial, contêm registos do

tempo decorrido entre um instante inicial e a ocorrência do evento de interesse, também conhecido como falha, lidando em geral com dados censurados (*Serranho, 2015*).

De salientar que o estudo de fenómenos em que a variável principal é o tempo-até-ao-evento não é obviamente um assunto exclusivo do estudo da sobrevivência a tratamentos do cancro e das ciências médicas, tendo aplicações variadas a diversas áreas da atividade humana, onde podemos destacar exemplos como:

1. Marketing/Gestão: tempo até ao abandono de um consumo a partir da subscrição de um serviço;
2. Psicologia Clínica: reincidência de consumo na sequência de programas de tratamento da toxicod dependência;
3. Engenharia: tempo até ocorrência da avaria de um componente;
4. Desporto: tempo de vida de um record Olímpico, tempo até desclassificação do batedor no jogo de Críquete.

## 2.1 Conceitos de Análise de Sobrevivência

Nesta secção vamos proceder à apresentação dos principais conceitos e variáveis em uso no âmbito da análise de sobrevivência, constituindo estes os elementos fundamentais do desenvolvimento das técnicas desta área da estatística.

### 2.1.1 Censura e Truncatura

Como referimos anteriormente, os dados de análise de sobrevivência são geralmente dados censurados, entendendo-se censura, neste domínio, como a impossibilidade de observar diretamente a ocorrência do evento de interesse e por esse motivo determinar com o máximo rigor de medida os "tempos de vida" dos indivíduos (*Rocha e Papoila, 2008*).

Saliente-se, contudo, que ao denominar de "censurados" esta designação não significa que estes dados não sejam incluídos no cálculo dos estimadores e modelos de regressão, mas antes que as técnicas aplicadas em análise de sobrevivência têm que ter em consideração esta característica específica dos dados.

Dentro da tipologia de dados censurados podemos definir tanto censura à direita, como censura à esquerda e, também, censura intervalar.

Em Análise de Sobrevivência, a censura à direita está relacionada com dados em que o tempo de vida excede o tempo de observação, havendo assim apenas um tempo de censura observado, que se sabe ser inferior ao tempo de vida (*Rocha e Papoila, 2008*).

A censura à esquerda, de forma similar, relaciona-se com dados referentes a indivíduos com um tempo de vida inferior a um determinado valor, nomeadamente, quando o evento ocorreu antes do indivíduo entrar em observação (*Rocha e Papoila, 2008*).

A censura intervalar está associada aos dados de indivíduos que têm tempos de observação fora de determinados intervalos de tempo ocorrendo, por exemplo, em estudos

clínicos longitudinais quando se consideram períodos de "follow-up" e existem indivíduos com formas irregulares de presença em diversos períodos de observação (*Rocha e Papoila, 2008*).

Um conceito diverso de censura e também usado em Análise de Sobrevida é o de truncatura, que corresponde a um processo de seleção de dados inerente à organização do estudo em que são estudados apenas indivíduos com determinada característica, sendo que esta poderá ser alterada ao longo do tempo, isto é, determinados indivíduos cuja característica venha a alterar-se poderão ser ocultadas dos procedimentos estatísticos da Análise de Sobrevida (*Rocha e Papoila, 2008*).

Salientar que, ao contrário do que acontece com os dados censurados, no caso de dados truncados as técnicas de Análise de Sobrevida não fazem inclusão destes em qualquer cálculo de estimadores ou modelos de regressão.

### 2.1.2 Variável Principal e Evento de Interesse

Em essência, o principal objeto de estudo da Análise de Sobrevida é uma variável aleatória contínua que se constitui como a medida do tempo desde um instante inicial determinado e até à ocorrência do evento de interesse. A esta variável é atribuída a designação de "tempo de sobrevivência" ou "tempo de vida" (*Serranho, 2015*).

Autores como (*Klein et al., 2013*) preferem considerar antes, em face da existência de dados censurados à direita, isto é, o tempo de vida sabe-se ser igual ou superior ao tempo de observação, duas variáveis que são independentes e condicionadas por outras covariáveis, descritas por

$$\begin{aligned} T_{surv} &: \text{ tempo de sobrevivência} \\ T_{cens} &: \text{ tempo de censura} \end{aligned}$$

Aqueles mesmos autores acrescentam que em termos práticos o **tempo de sobrevivência**  $T$  observado é o mais pequeno de

$$T = \min(T_{surv}, T_{cens})$$

que ademais poderá ser sumariado no par  $(T, D)$  em que  $D$  é o indicador que informa sobre se  $T$  foi observado ou censurado.

Esclarecido, neste âmbito, o conceito de tempo de sobrevivência, acrescenta-se que este está associado à noção de probabilidade de sobrevivência que se obterá por via da função de sobrevivência  $S(t)$  em geral formulada com a expressão

$$S(t) = P(T > t), t \geq 0 \tag{2.1}$$

em que  $t$  indica o instante para o qual se define a probabilidade e  $T$  a variável aleatória contínua que é o tempo de sobrevivência.

### 2.1.3 Covariáveis

Para além do tempo de sobrevivência, é evidente que poderão estar associados outros fatores que potenciam ou inibem a ocorrência do evento e que se designam

como covariáveis (*Serranho, 2015*).

Estas covariáveis podem manter o valor idêntico para o mesmo indivíduo ao longo do tempo de observação (i.e. podem ser constantes para cada indivíduo) ou, ao invés, manifestarem valores diferentes durante a observação, dizendo-se estas últimas "covariáveis dependentes do tempo".

De salientar que ao inserirmos estas covariáveis nos modelos de sobrevivência estamos a recolher mais informação sobre a forma como se comporta a probabilidade de sobrevivência na presença de cada um e do conjunto dos fatores que influenciam o tempo de vida.

## 2.2 Estimadores Não Paramétricos da Função de Sobrevivência

Nesta secção vamos explicitar melhor o que é considerado um dos principais instrumentos de caracterização do "tempo de sobrevivência" associado a um evento de interesse e referente a um grupo de indivíduos.

Diversos investigadores propuseram metodologias que permitissem uma melhor caracterização e estimação da função de sobrevivência.

Com os dados plasmados na variável *tempo de sobrevivência*, o intuito seguinte do investigador será identificar a probabilidade de sobrevivência que será obtida através da função de sobrevivência tal como indicada na expressão 2.1.

Reiterando a função de sobrevivência será a função que faz corresponder a um instante  $t$  a correspondente probabilidade de um indivíduo sobreviver para além desse instante.

Relembremos que com dados censurados está inibido o recurso à metodologia estatística convencional, baseada unicamente nos indivíduos em observação e nos tempos explicitamente registados do evento de interesse.

Daí a importância da proposta de Kaplan e Meier (*Kaplan e Meier, 1958*) para a definição de um estimador não paramétrico da função de sobrevivência que se designa atualmente por estimador de Kaplan-Meier e que passamos a descrever sucintamente nos parágrafos seguintes.

Seguindo o raciocínio de (*Klein et al., 2013*) e tendo como ponto de partida uma amostra de  $n$  observações independentes

$$(t_1, d_1), (t_2, d_2), \dots, (t_n, d_n) \in (T, D) \quad (2.2)$$

com  $d_i = 0$  se  $t_i$  corresponde a tempo de censura e  $d_i = 1$  se  $t_i$  corresponde a um tempo de falha.

Notamos o conceito de conjunto de risco, que é composto pelos indivíduos em risco num determinado instante e que inclui todos os que sobreviveram até esse instante mais os que, por via de dados censurados, se sabe terem tempo de vida superior a esse instante.

Considerando assim o conjunto de risco e tendo a informação referente ao registo do número de eventos ocorridos nesse instante, o estimador de Kaplan-Meier é obtido por via da expressão

$$\hat{S}(t) = \prod_{s \leq t} \frac{n_s - e_s}{n_s} \quad (2.3)$$

em que  $n_s$  corresponde à cardinalidade do conjunto de risco no instante  $s$  e  $e_s$  o número de eventos ocorridos nesse instante (Serranho, 2015).

Pode-se demonstrar que a distribuição do estimador de Kaplan-Meier converge assintoticamente para a distribuição normal e que é um estimador consistente de  $S(t)$  (Serranho, 2015).

Com este estimador podemos deduzir o intervalo de confiança para o verdadeiro valor da função de sobrevivência num determinado instante  $t_i$  que será obtido, como referem (Rocha e Papoila, 2008), pela denominada Fórmula de Greenwood

$$\left( \hat{S}(t_0) - z_{1-\alpha/2} \sqrt{\hat{v}ar \hat{S}(t_0)}, \hat{S}(t_0) + z_{1-\alpha/2} \sqrt{\hat{v}ar \hat{S}(t_0)} \right) \quad (2.4)$$

Dado que é um intervalo simétrico pode acontecer que os seus limites estejam fora do intervalo  $(0, 1)$  para estimativas de  $S(t)$  próximas de 0 ou 1.

Um procedimento possível será como propuseram Kalbfleish e Prentice (Kalbfleish e Prentice, 2002) definir de uma nova variável com base na expressão

$$\hat{U}(t) = \ln(-\ln[\hat{S}(t)]) \quad (2.5)$$

Pode-se, pois demonstrar, (Serranho, 2015) que a variância do estimador  $\hat{U}$  é dada por

$$\hat{V}(\hat{U}(t)) = \frac{\sum_{i:t_i < t} \frac{f_i}{n_i(n_i - f_i)}}{(\ln[S(t)])^2} \quad (2.6)$$

e assim o intervalo de confiança de  $S(t)$  é dado por

$$IC_{1-\alpha}(t) = \left[ \hat{S}(t)^{\exp(z_{\alpha/2} \cdot \sqrt{\hat{V}(\hat{U}(t))})}, \hat{S}(t)^{\exp(z_{1-\alpha/2} \cdot \sqrt{\hat{V}(\hat{U}(t))})} \right] \quad (2.7)$$

sendo que este intervalo tem sempre limites entre 0 e 1 pois a estimativa  $\hat{S}(t)$  tem sempre valores entre 0 e 1.

Se, como referem Klein et al. (Klein et al., 2013), pressupusermos que a função de sobrevivência é uma função diferenciável, com uma variável que pertence a uma distribuição contínua, definimos um conceito de "função taxa de falha" ou "função de risco" como a que corresponde à razão entre a probabilidade de um evento ocorrer num determinado intervalo de tempo e a amplitude desse intervalo, condicionada ao facto de o evento não ter ocorrido antes e que se pode ser representada como

$$\lambda(t) = \lim_{h \rightarrow 0} \frac{P(t \leq T \leq t+h | t \leq T)}{h} = \lim_{h \rightarrow 0} -\frac{S(t+h)}{S(t)} \quad (2.8)$$

Trata-se de um indicador da variação instantânea da sobrevivência, veja-se, a título de exemplo, (Serranho, 2015), tendo um conceito a que está ligado diretamente e que se

denomina de função de "taxa falha acumulada" ou "taxa de risco cumulativo" definido através de

$$\Lambda(t) = \int_0^t \lambda(s) ds \quad (2.9)$$

em que  $\Lambda(t)$  corresponde ao somatório de probabilidades instantâneas de "morte", ou de ocorrência do evento, desde o instante inicial até ao correspondente instante  $t$ , indicando-nos a probabilidade global ou acumulada de "morte" nesse intervalo de tempo  $[0, t]$ .

As funções de sobrevivência e taxa de falha (*Serranho, 2015*) têm uma relação muito próxima que se pode apresentar como

$$\Lambda(t) = -\ln(S(t)) \quad (2.10)$$

$$S(t) = \exp(-\Lambda(t)) \quad (2.11)$$

Saliente-se, como referem (*Klein et al., 2013*), que a função taxa de falha será bem definida se a função de sobrevivência for diferenciável, o que não é líquido no caso do estimador de Kaplan-Meier, motivo pelo qual será melhor estimar a função taxa de falha acumulada  $\Lambda(t)$ .

Uma alternativa ao estimador de Kaplan-Meier será usar o estimador de Nelson-Aalen (*Serranho, 2015*), definido como

$$\tilde{H}(t) = \sum_{s \leq t} \frac{e_s}{n_s} \quad (2.12)$$

em que  $\tilde{H}(t)$  representa o estimador da função de risco acumulado, também conhecida como função taxa de falha acumulada empírica, sendo evidente que se pode obter a partir desta um estimador da função de sobrevivência, também conhecido (*Rocha e Papoila, 2008*) como estimador de Breslow, por via da expressão

$$\tilde{S}(t) = \exp(-\tilde{H}(t)) = \prod_{s \leq t} \exp\left(-\frac{e_s}{n_s}\right) \quad (2.13)$$

Saliente-se que, para pequenas amostras, o estimador de Nelson-Aalen apresenta um melhor comportamento do que o estimador de Kaplan-Meier, pois

$$e^{-x} \approx 1 - x$$

tendo aqui

$$\exp(-e_i/n_i) \approx 1 - (d_i/n_i)$$

revelando que as estimativas são muito aproximadas para amostras de dimensão elevada (*Serranho, 2015*).

## 2.2.1 Comparação de Funções de Sobrevivência

Uma vez identificado um estimador da função de sobrevivência e obtida uma estimativa da função de sobrevivência, a análise dos dados pode concretizar-se num outro objetivo que corresponde à comparação da distribuição do tempo de vida de diversos grupos de indivíduos, determinando se existem diferenças estatisticamente significativas entre as funções de sobrevivência desses grupos (*Serranho, 2015*).

Se, por um lado, se pode proceder à comparação através da representação gráfica, quer da função de risco cumulativo, quer da função de sobrevivência, podendo discriminar visualmente as diferenças entre funções de sobrevivência, uma forma mais rigorosa consiste na avaliação da existência de diferenças significativas entre as várias curvas o que implica a aplicação de um ou mais testes de comparação de curvas de sobrevivência, neste caso testes não paramétricos (*Serranho, 2015*).

Estes testes destinam-se, em geral, a validar as hipóteses estabelecidas como se indica em seguida

$$\left\{ \begin{array}{l} H_0: S_1(t) = S_2(t) = \dots = S_p(t) \\ H_1: S_i(t) \neq S_j(t) \quad \exists i, j : i \neq j \end{array} \right\} \quad (2.14)$$

O teste mais usado no domínio da análise de sobrevivência é o teste de *Log-Rank*.

Seguindo (*Rocha e Papoila, 2008*) para a construção do teste de *Log-Rank* parte-se da listagem ordenada de todos os tempos dos eventos em todos os grupos

$$t_1 < t_2 < \dots < t_m$$

em seguida podemos definir uma tabela de contingência com a informação relevante em cada instante  $t_k$  (vd. tabela 2.1).

Tabela 2.1: Tabela de Frequências para o Teste de Log-Rank

	Grupo 1	Grupo 2	...	Grupo m	Total
Evento	$e_{1k}$	$e_{2k}$	...	$e_{mk}$	$e_k$
Não-Evento	$n_{1k} - e_{1k}$	$n_{2k} - e_{2k}$	...	$n_{mk} - e_{mk}$	$n_k - m_k$
Total	$n_{1k}$	$n_{2k}$	...	$n_{mk}$	$n_k$

Correspondendo aqui  $e_{ik}$  ao número de eventos do grupo  $i$  no instante  $k$  e  $n_{ik}$  ao número de indivíduos do grupo de risco  $i$  no instante  $k$ .

Para a aplicação do teste de Log-Rank define-se a hipótese nula que assume a igualdade das funções de sobrevivência podendo considerar-se a distribuição condicional das frequências observadas em cada célula da tabela, dados os totais marginais. Isto é, considera-se a distribuição de frequência de apenas uma célula da tabela, pois as outras frequências são determinadas pelos totais marginais fixos.

O que significa que se pode dessa forma proceder ao cálculo da distribuição condicional das frequências observadas.

Assim a probabilidade de  $d_{1k}$  condicional aos valores marginais será dada por:

$$p(e_{1k}|e_k, n_k) = \frac{\binom{e_k}{e_{1k}} \cdot \binom{n_k - e_k}{n_{1k} - e_{1k}}}{\binom{n_k}{n_{1k}}}$$

o que corresponde a uma distribuição hipergeométrica multivariada, cujo valor médio e variância permitem obter uma medida global do desvio dos valores observados em relação aos valores esperados e a construção de uma estatística de teste obtida com

$$T = \frac{e' \times e}{V} \sim \chi_{m-1}^2$$

em que  $e$  corresponde ao vetor com todos os instantes  $k$  que contêm o somatório entre o valor observado e o valor esperado para o número de falhas dado, e o valor de referência vem dado pela distribuição  $\chi^2$  com graus de liberdade igual a número de grupos menos 1.

## 2.3 Modelo de Cox

Em 1972 David Cox apresentou um modelo que "rapidamente se tornou no modelo de regressão mais utilizado na análise de tempos de sobrevivência" (*Rocha e Papoila, 2008*). Trata-se de um modelo de regressão semi-paramétrico, porque combina uma componente não-paramétrica e uma parte com os usuais parâmetros de regressão (*Serranho, 2015*).

Nesta secção vamos apresentar as principais características do modelo de Cox, relevando a importância no âmbito da análise de sobrevivência.

Por comparação com as abordagens não paramétricas, este modelo permite incluir informação com análise de covariáveis contínuas e da influência destas no comportamento da função de sobrevivência.

Contudo, o ajustamento deste modelo aos dados pressupõe previamente a verificação da hipótese de riscos proporcionais, isto é, que as funções de risco correspondentes a dois indivíduos, com um valor diferente de uma covariável, são proporcionais.

### 2.3.1 Componentes do Modelo de Cox

De uma forma expedita podemos dizer que, para um conjunto de covariáveis  $x_1, x_2, \dots, x_m$ , este modelo pressupõe que a taxa de falha será dada por

$$\lambda(t|x_1, x_2, \dots, x_m) = \lambda_0(t) \cdot \exp\{\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m\} \quad (2.15)$$

sendo  $\lambda_0(t)$  a taxa de falha basal, i.e., quando todas as covariáveis são nulas.

Na construção teórica do modelo para obter a estimativa da função taxa de falha, ou função de risco, começa-se por estimar os parâmetros  $\beta_1, \beta_2, \dots, \beta_m$  e posteriormente a função  $\lambda_0(t)$  - taxa de falha basal ou função de risco basal, terminando assim por obter a estimativa da função de sobrevivência.

Para obter a componente paramétrica inicia-se com a função de verosimilhança



$$\mathcal{L}(t_i, \theta) = \prod_{i=1}^n [\lambda(t_i, \beta)]^{\delta_i} \cdot S(t_i, \beta) \quad (2.16)$$

que pode também ser expressa como

$$\mathcal{L}(t_i, \theta) = \prod_{i=1}^n [\lambda_0(t_i)]^{\delta_i} \cdot \exp\{\delta_i(\beta_1 x_1 + \dots + \beta_m x_m)\} \cdot [S_0(t_i) \cdot \exp \beta_1 x_1 + \dots + \beta_m x_m] \quad (2.17)$$

tendo Cox usado o método da máxima verosimilhança parcial com o fim de permitir a realização de inferência na presença de parâmetros perturbadores, entendendo-se que  $\lambda_0(t)$  será uma função perturbadora (*Rocha e Papoila, 2008*).

Dessa forma, considerando a probabilidade condicionada

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (2.18)$$

que representa a probabilidade de um sujeito ter um evento - (A) - no instante  $t_i$  condicionada ao facto de existir uma falha em  $t_i$  de entre os indivíduos que não verificaram um evento nesse instante  $t_i$  - B.

Substituindo, para cálculo, a probabilidade condicionada referida em cima vem dada por

$$P_i = \frac{\lambda_0(t) \cdot \exp\{\beta_1 x_1^{(j)} + \dots + \beta_m x_m^{(j)}\}}{\lambda_0(t) \cdot \sum_{k \in R_i} \exp\{\beta_1 x_1^{(k)} + \dots + \beta_m x_m^{(k)}\}} \quad (2.19)$$

a razão de probabilidades de falha do lado direito anula o elemento  $\lambda_0$  que, numa primeira fase, é desconhecido.

Com a equação anterior estabelecemos o objetivo de maximizar o produto das  $P_i$  para o instante de falha  $t_i$  o que se pode obter por via da função de verosimilhança parcial (*Serranho, 2015*) dada por

$$\mathcal{L}_P = \prod_{i=1}^n \left( \frac{\exp\{\beta_1 x_1^{(j)} + \dots + \beta_m x_m^{(j)}\}}{\sum_{k \in R_i} \exp\{\beta_1 x_1^{(k)} + \dots + \beta_m x_m^{(k)}\}} \right) \quad (2.20)$$

Aplicando métodos numéricos podemos identificar a estimativa dos coeficientes  $\beta_i$  e prosseguir para o cálculo da estimativa da componente não paramétrica (*Rocha e Papoila, 2008*).

A parte não paramétrica do modelo corresponde à taxa de falha basal  $\lambda_0(t)$  sendo esta estimada por via do estimador de Breslow calculado com

$$\hat{\Lambda}_0(t) = \sum_{i:t_i < t} \frac{e_i}{\exp(\beta_1 x_1^{(k)} + \dots + \beta_m x_m^{(k)})} \quad (2.21)$$

existindo, para o mesmo estimador, uma outra alternativa proposta por Kalbfleisch et al. (*Kalbfleish e Prentice, 2002*).

### 2.3.2 Testes da Hipótese de Riscos Proporcionais

Como referimos na secção anterior, a construção do modelo de Cox parte da construção de um modelo que pressupõe que a taxa de falha, ou função de risco, tem uma relação proporcional que se pode definir como

$$\lambda_i(t) = \lambda_0(t)c_i$$

sendo que neste caso  $c_i = \exp(X_i^T \beta)$  (Klein et al., 2013).

Importa assim testar previamente ao ajustamento dos dados se, de facto, se verifica a hipótese de os coeficientes serem independentes do tempo contra a hipótese de os coeficientes não serem constantes ao longo do tempo, sendo óbvio que neste último caso o valor da função de risco teria variações em função do tempo e os riscos não seriam proporcionais.

Para esse efeito o usual é utilizar os resíduos padronizados de Schoenfeld para aferir da hipótese de riscos proporcionais (Serranho, 2015).

Com a presença de dados censurados a definição e cálculo de resíduos carece de maiores cuidados do que no caso das metodologias estatísticas convencionais, por esse motivo várias propostas existem para cálculo de resíduos para o modelo de Cox (Rocha e Papoila, 2008).

Com aquela motivação Cox e Snell, em 1968, propuseram a definição de um tipo de resíduo baseado numa amostra proveniente de determinada distribuição conhecida (Cox e Snell, 1968).

Assim o resíduo do  $i$ -ésimo indivíduo será obtido como  $r_i = \hat{H}(t_i) = \exp(\hat{\beta}'z_i)\hat{H}_0(t_i)$  e se o modelo é ajustado então os valores estimados  $\hat{H}(t_i)$  serão semelhantes aos valores observados  $H(t_i)$ .

Por isso os  $r_i$  terão comportamento aproximado com uma amostra aleatória proveniente de uma população com distribuição exponencial com a expressão  $\text{Exp}(1)$ .

No caso dos dados censurados os resíduos de Cox-Snell são modificados de forma a ser adicionada uma constante positiva que respeite a distribuição exponencial da amostra aleatória para o tempo de vida  $T \sim \text{Exp}(1)$ .

A constante terá valor  $c = 1$  e os resíduos modificados de Cox-Snell vêm dados como

$$\begin{cases} r_i = r_i & \text{para } t_i \text{ de evento observado} \\ r_i = r_i + 1 & \text{para } t_i \text{ de evento censurado} \end{cases} \quad (2.22)$$

a principal desvantagem deste modelo é que estes resíduos não medem o afastamento do modelo, e, por outro lado, estes resíduos não conseguem distinguir entre uma errada adequação do modelo e a verificação do pressuposto de riscos proporcionais (Serranho, 2015).

Outra proposta de definição de resíduos para o modelo de Cox foi a adiantada por Schoenfeld em 1982 (Schoenfeld, 1982), que difere dos resíduos de Cox-Snell (Rocha e Papoila, 2008) por dois motivos:

- não ser necessário ter uma estimativa da função de sobrevivência;

- a cada indivíduo corresponder um conjunto de valores, sendo um por cada covariável incluída no modelo de regressão de Cox.

O cálculo dos resíduos de Schoenfeld dos indivíduos do conjunto de risco define-se (*Klein et al.*, 2013) com a expressão

$$score(t) = x_{ev}(t) - \bar{x}(\hat{\beta}, t) \quad (2.23)$$

em que  $x_{ev}(t)$  é o valor médio dos indivíduos que tiveram uma falha no instante  $t$  ou seja

$$x_{ev}(t) = \sum_i x_i \cdot \Delta N_i(t) / \Delta \bar{N}(t) \quad (2.24)$$

Grambsch e Therneau (*Grambsch e Therneau*, 1994) propõem que seja desenhado o gráfico dos componentes dos denominados "resíduos padronizados de Schoenfeld" obtidos com

$$score * (t) = \hat{\beta} + V(t)^{-1} score(t) \quad (2.25)$$

em que  $V(t) = var(x|\hat{\beta}, t)$  é a covariância ponderada obtida para o conjunto de risco (*Klein et al.*, 2013).

Dessa forma, quando estamos a testar a hipótese nula  $\beta(t) = \beta$  pretendemos verificar se o gráfico dos referidos resíduos padronizados em relação a  $t$  corresponde a uma linha horizontal (*Serranho*, 2015).

### 2.3.3 Seleção de Variáveis do Modelo

A inclusão de variáveis no modelo tem como principal objetivo associar todas as que podem ser identificadas como as que têm uma importância e influência significativa na expressão da função de sobrevivência (*Serranho*, 2015).

Segundo (*Turkman e Silva*, 2000) é frequente a ocorrência de um elevado número de covariáveis que podem ser potencialmente importantes para explicar a variabilidade de dados.

Dado que no processo de seleção existe uma série de modelos em consideração, podem descrever-se vários tipos que são comumente referidos durante o mesmo processo:

1. o modelo completo ou saturado, é o modelo linear generalizado, sendo considerado por ser necessário na formulação da teoria de seleção de modelos;
2. o modelo nulo, com um único parâmetro, que corresponde a assumir que todas as variáveis têm o mesmo valor médio;
3. o modelo maximal, é o modelo que contém o maior número de parâmetros sendo, portanto, o mais complexo;
4. o modelo minimal, que é o modelo com menor número de parâmetros que ainda se ajustam aos dados, embora adaptado aos dados e podendo até ser adequado para réplicas do estudo, pode esconder características ainda importantes dos dados.

A questão reside, em geral, em decidir qual o equilíbrio entre o efeito de inclusão ou exclusão de um termo na discrepância entre  $y$  e  $\hat{\mu}$  e a correspondente complexidade de um modelo maior (*Turkman e Silva, 2000*).

Há, pois, necessidade de estabelecer uma estratégia para a seleção do melhor modelo (*Turkman e Silva, 2000*). Estabelecendo os termos gerais, podemos dizer que se tivermos um modelo  $M_1$  de um modelo  $M$ , por eliminação da variável  $x_i$ , importa saber da adequabilidade do submodelo testando formalmente

$$H_0 : \beta_i = 0 \text{ vs. } H_1 : \beta_i \neq 0 \quad (2.26)$$

para esse efeito podemos considerar as seguintes estatísticas de teste:

- teste da razão de verossimilhanças,
- teste de Wald,
- teste de Rao.

Segundo Fahrmeier e Tritz, citados por (*Turkman e Silva, 2000*) é preferível em geral usar o teste da razão de verossimilhanças se o número de covariáveis for pequeno e as amostras tiverem uma dimensão moderada.

Em geral pretende-se uma economia de covariáveis em que se sinalizam as variáveis significativas e se retiram as restantes (*Serranho, 2015*).

Com efeito, as covariáveis não significativas podem causar efeitos perniciosos na modelação pelo que deverão ser ignoradas na análise dos dados.

Uma forma de o fazer é considerar o teste de hipóteses definido como em 2.22. contrapondo a estatística baseada no teste de Wald ou no teste da razão de verossimilhanças.

Para o efeito recorre-se ao "método stepwise", introduzindo ou eliminando as variáveis passo-a-passo (*Serranho, 2015*).

Assim, pode-se por um lado partir do conjunto total das variáveis, retirando uma a uma e testando qual delas menos afeta o modelo, continuando-se até que qualquer das variáveis tenha um efeito significativo no modelo.

No caso em que menos se altera, retira-se a covariável correspondente e continua-se o processo retirando, então, uma a uma, as variáveis que permaneceram no modelo e voltando a retirar as que menos influência têm no modelo.

Quando ao retirar qualquer uma das variáveis se verifica uma alteração significativa no modelo, expresso no teste de razão de verossimilhanças, o processo é parado.

Desta forma eliminam-se variáveis que possam estar correlacionadas e que, portanto, tenham efeito duplicado no modelo (*Serranho, 2015*).

## 2.4 Aplicações da Análise de Sobrevivência no Âmbito de Fenómenos Desportivos

Nesta secção vamos apresentar alguns exemplos de aplicação de métodos da análise de sobrevivência a fenómenos do âmbito das atividades desportivas revendo a literatura que foi aplicada neste domínio.

Um exemplo importante, em relação ao qual pretendemos estender os seus objetivos com a realização desta dissertação, correspondeu a um estudo elaborado por D. Nevo e Y. Ritov (*Nevo e Ritov, 2012*). Trata-se de um estudo que abordou a ocorrência de eventos de golo na liga inglesa durante duas épocas desportivas e cujo principal objetivo era identificar relações entre o tempo do primeiro golo e o tempo do segundo golo. Os autores concluíram que o tempo do primeiro golo tanto poderia apressar como retardar a ocorrência do segundo golo, dependendo do momento em que tinha sido marcado o segundo golo. Em suma, o modelo que definiram como mais ajustado para os dados continha 5 variáveis: as *odds* da equipa de casa, a época, uma variável dicotómica relativa a ocorrência de golo, o tempo do primeiro golo e o logaritmo do tempo a partir do primeiro golo.

Outro exemplo de aplicações em fenómenos desportivos constitui o estudo sobre os records de Jogos Olímpicos (*Hollifield e Zarn, 2012*), os autores definiram como evento de interesse a concretização de uma nova marca máxima de jogos olímpicos pretendendo ter um melhor conhecimento do que determinará a duração dos records olímpicos, para além de usar o estimador de Kaplan-Meier e de extensões do modelo de Cox. Os autores, usando os conceitos de eventos recorrentes aplicaram dois estimadores diferentes tendo também ajustado um modelo logístico de tempo discreto para eventos repetidos. Por fim concluíram que dos 51 eventos competitivos em análise foram estimados que entre 20.12 e 31.14 records olímpicos seriam derrubados.

Ainda um outro exemplo de aplicações em fenómenos desportivos vem da análise de tempo de jogo de batedores de Críquete (*Stevenson e Brewer, 2017*) tendo os autores recorrido a métodos da estatística bayesiana para especificar os modelos de análise dos jogadores. Em conclusão os autores referem previsões das habilidades dos batedores no início da época desportiva que se seguia na Nova Zelândia.

Para terminar, um outro exemplo consistiu na aplicação do estimador de Kaplan-Meier em conjunto com o teste de Log-Rank para obter uma estimativa da permanência dos participantes nas atividades físicas de organização municipal (*Monteiro e Silva, 2014*).



# Capítulo 3

## Metodologia

### 3.1 Metodologia de Investigação

Tendo estabelecido como principal objetivo deste estudo a caracterização e comparação de funções de sobrevivência de eventos de golo de jogos de futebol das principais ligas europeias profissionais importa esclarecer a metodologia de investigação adotada e as principais escolhas e passos concretizados com vista a atingir os objetivos definidos: identificação e análise de covariáveis, efeitos de as covariáveis na função de sobrevivência e comparação de funções de risco de eventos diferentes tendo por base os conceitos e métodos apresentados genericamente no capítulo anterior.

No presente capítulo damos então conta da metodologia de recolha de dados, indicaremos também como foram estruturados e armazenados os dados, descreveremos brevemente o software estatístico utilizado na análise de dados e vamos listar as principais etapas da metodologia de análise estatística implementada.

### 3.2 Metodologia de Recolha de Dados

A recolha de dados referentes aos registos dos eventos de golo teve como fundamento o conhecimento prévio da disponibilidade de páginas "web", integradas em portais, que disponibilizam informação sobre os resultados de jogos de futebol das competições de topo de diversos países do mundo.

Referimo-nos nomeadamente aos portais presentes nos endereços de internet:

1. [www.scorespro.com/soccer](http://www.scorespro.com/soccer),
2. [www.football-data.co.uk/data.php](http://www.football-data.co.uk/data.php),
3. [www.oddsportal.com/soccer](http://www.oddsportal.com/soccer),

todos estes acedidos em junho de 2017.

Para o efeito considerámos o uso da técnica de "web scraping" que, de forma automatizada, permitiu a recolha de dados.

Dessa forma, usando o pacote Beautiful Soup, da linguagem Python, criámos rotinas que permitiram a navegação pelas etiquetas ("tags") das páginas em HTML, que continham as informações relativas a:

1. listagem de desafios por jornada;
2. registo de informações sobre cada desafio;
3. dados relativos a *odds* definidas pelas casas de apostas.

Para o item referentes às *odds* usámos o pacote "pandas", da linguagem Python, para importar um ficheiro, com dados em formato CSV, e desenvolvemos uma rotina para o armazenamento direto dos mesmos em base de dados.

### 3.3 Armazenamento de Dados

Num primeiro passo procedemos ao desenho de um esquema de uma base de dados relacional para identificar a estruturas de dados a integrar na base de dados.

O uso do motor de base de dados SQLite, versão 3, e o recurso à ferramenta de interface gráfico "DB SQLite Browser", permitiu-nos uma consulta expedita dos dados e do respetivo esquema, bem como a realização de testes experimentais com comandos da linguagem SQL, principalmente consultas de dados.

Para além das tabelas, que armazenaram de forma persistente os dados, criámos "vistas", i.e. conjuntos que são resultados de consultas pré-definidas, o que permitiu a construção das denominadas *dataframes* (quadros de dados) que foram usadas de forma mais ou menos direta nos diversos procedimentos da Análise de Sobrevida.

### 3.4 Software Estatístico

Para a realização das análises estatísticas escolhemos o software da linguagem R, incluindo diversos pacotes principalmente os usados em estudos de análise de sobrevivência: "survival", "survcomp", "cumhaz".

Por fim, para efeitos de registo e melhor manipulação das análises de dados recorremos à ferramenta "Jupyter Notebooks", o que nos permitiu compilar documentos com código, equações, visualizações e texto descritivo diverso (vd. *Jupyter homepage*).

Por outro lado, recorreremos a pacotes de software auxiliar como as bibliotecas da linguagem R: "RSQLite", "ggplot2" e "colors", que asseguraram a realização de tarefas intermédias da análise estatística: conexão com a base de dados, desenho de gráficos e uso de cores.

### 3.5 Análise Estatística

Na secção anterior descrevemos o que constituiu a "caixa de ferramentas computacionais", pelo que importa agora dar conta da metodologia de análise estatística descrevendo o conjunto de fases em que se desenrolou o processo de tratamento e análise de dados.



### 3.5.1 Eventos de Interesse

Com os dados recolhidos definimos o propósito de estabelecer o que entendíamos como eventos de interesse com relevância para este estudo. Assim, considerando que já existia um estudo que tinha analisado os eventos referentes ao primeiro e ao segundo golo do jogo, definimos como eventos de interesse, para além daqueles dois (primeiro e segundo golo do jogo) também o terceiro golo do jogo.

Importa salientar que a sequência de procedimentos de análise estatística foi repetida para todos os eventos de interesse, pelo que, para além dos eventos referidos em cima, também considerámos os seguintes, referentes ao segundo e terceiro golos do jogo:

1. segundo golo:
  - golo de 1-0 para 2-0,
  - golo de 0-1 para 0-2,
  - golo de 1-0 para 1-1,
  - golo de 0-1 para 1-1;
2. terceiro golo:
  - golo de 1-1 para 2-1,
  - golo de 1-1 para 1-2.

O nosso estudo debruçou-se então, inicialmente, sobre o evento de interesse que foi o primeiro golo do jogo, estabelecendo desde logo uma distinção entre dois tipos de eventos:

- primeiro golo foi marcado pela equipa de casa;
- primeiro golo foi marcado pela equipa de fora (visitante).

Por outro lado, a título suplementar, estipulámos dois outros tipos de eventos de interesse que correspondiam também ao primeiro golo do jogo mas tendo como referência a equipa que marcou esse golo, assim definimos:

- equipa de casa marca o seu primeiro golo;
- equipa de fora (visitante) marca o seu primeiro golo.

### 3.5.2 Estimador da Função de Sobrevida - Kaplan-Meier

Para o estudo das funções de sobrevivência associadas aos eventos iniciámos com a obtenção da respetiva estimativa de Kaplan-Meier, pois entendemos ser este um estimador adequado para o efeito, quer por ser não paramétrico, quer por não ter pressupostos de aplicação e não ser parte de qualquer condicionante do comportamento daquela função.

Uma vez obtida a referida estimativa aplicámos o teste de Log-Rank com vista a comparar as diversas funções de sobrevivência estimadas e assim conhecer da existência eventual de diferenças significativas entre os diversos grupos de jogos.

### 3.5.3 Modelo de Cox

A identificação, apresentação gráfica e comparação de funções de sobrevivência por via do estimador de Kaplan-Meier, foi seguida pelo estudo do ajustamento do modelo de Cox aos dados.

Para este efeito considerámos inicialmente as covariáveis disponíveis e procedemos à análise dos resíduos padronizados de Schoenfeld, o que nos permitiu identificar as covariáveis que comprovadamente respeitavam a hipótese de riscos proporcionais, sendo esta hipótese um pressuposto fundamental do modelo de Cox.

Identificadas as covariáveis que respeitavam a hipótese de riscos proporcionais, passámos a uma segunda fase do ajustamento do modelo de Cox em que se concretizou um processo de seleção de variáveis usando o método "step wise" (passo a passo).

Em resumo, partindo do que é referido por (*Turkman e Silva, 2000*) como "modelo maximal", em que foram incluídas todas as covariáveis, retirámos, uma a uma, as que se identificaram como menos significativas.

Assim, podemos descrever o método "step wise" da seguinte forma:

1. retirou-se a variável do modelo;
2. aplicou-se o teste de razão de verosimilhanças (TRV), entre o modelo atual e o modelo anterior;
3. verificou-se se o valor da razão de verosimilhanças era significativa;
4. continuou-se o processo no caso da TRV não ser significativa, parando-se o processo caso contrário.

Nos casos em que se parou o processo foi incluída a última variável retirada pelo procedimento.

# Capítulo 4

## Apresentação de Resultados

Neste capítulo vamos apresentar os resultados obtidos na sequência das análises realizadas de acordo com a metodologia apresentada no capítulo 3. Uma breve anotação para esclarecer e dar conta da definição global dos eventos em estudo: elegemos os eventos de golo correspondentes ao 1.º, 2.º e 3.º golos dos jogos em análise. Iniciamos com uma secção referente ao 1.º golo da equipa (casa/fora) onde calculamos a estimativa de Kaplan-Meier da função de sobrevivência referente aos dois casos: 1.ª vez que a equipa de casa marca um golo e 1.ª vez que a equipa de fora marca um golo. Trata-se do caso de eventos em que, como se pode verificar, é fixada a equipa e, assim se pode dizer, identificado o evento em função desta.

Este foi um exercício que se constituiu como uma breve ilustração de eventos de golo que podem ser analisados para comparação com os outros eventos de golo que serão, geralmente, os mais "importantes" para a marcha até ao resultado final dos desafios em estudo.

Nas secções seguintes procedemos à análise dos eventos de golo identificando em geral a configuração do marcador a que estão associados.

Como se poderá constatar as combinações possíveis aumentam à medida que aumenta a ordem do evento de golo. Este facto é relevante para identificar a importância de alguns eventos para efeitos de comparação das respectivas funções taxa de risco.

Estas foram as principais razões que levaram a que, no caso do 3.º golo do jogo, a análise fosse limitada, para efeitos de comparação, a 2 eventos, para nós os mais relevantes, correspondentes ao evento que produziu uma configuração de desempate no marcador, nomeadamente:

1. de 1-1 para 2-1,
2. de 1-2 para 1-2.

### 4.1 1.º Golo da Equipa

Neste caso o evento de interesse correspondeu ao golo que a equipa (casa/fora) marcou pela primeira vez, sendo este registo considerado observado independentemente de terem ou não ocorrido antes golos da equipa adversária.

Em primeiro lugar considerou-se o evento correspondente ao 1.º golo marcado pela equipa de casa tendo sido censurados todos os registos relativos a jogos em que não ocorreu qualquer golo e os relativos a jogos em que só ocorreram golos da equipa de fora.

Em segundo lugar considerámos o evento correspondente ao 1.º golo da equipa de fora tendo a identificação de censuras sido semelhante ao descrito no parágrafo anterior.

Também foi considerada a variável “ligas” que permitiu identificar a que liga pertenciam os jogos.

#### 4.1.1 Estimativa de Kaplan-Meier da Função de Sobrevivência de Equipa de Casa Marca Respetivo 1.º Golo

Os quartis obtidos para este evento foram os indicados na tabela 4.1.

Tabela 4.1: Quartis de Tempo de Sobrevivência para Equipa de Casa Marca Respetivo 1.º Golo

Liga	1.º Quartil	Mediana	3.º Quartil
Portugal	22	49	ND
Espanha	21	45	90
Alemanha	19	43	85
Inglaterra	22	47	89
França	23	50	90
Itália	21	48	88

ND - não disponível

##### 4.1.1.1 Função de Sobrevivência

Para estes dados procedeu-se ao cálculo da estimativa de Kaplan-Meier da função de sobrevivência e à aplicação do Teste de Log-Rank.

O valor-p obtido foi de 0.0378 e o gráfico da função de sobrevivência obtida pela Estimador de Kaplan-Meier foi o apresentado na figura 4.1.

#### 4.1.2 Estimativa de Kaplan-Meier da Função de Sobrevivência de Equipa de Fora Marca Respetivo 1.º Golo

Os quartis obtidos para este evento foram os indicados na tabela 4.2.

##### 4.1.2.1 Função de Sobrevivência

Considerando estes dados procedeu-se ao ajustamento do Estimador de Kaplan-Meier e à aplicação do Teste de Log-Rank.

O valor-p obtido foi de 0.000284 e o gráfico da função de sobrevivência obtida pela Estimador de Kaplan-Meier está apresentado na figura 4.2.

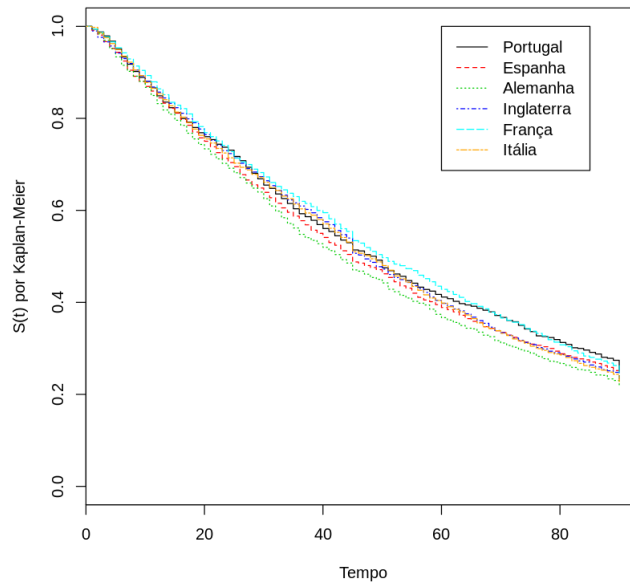


Figura 4.1: Estimativa de Kaplan-Meier da Função de Sobrevivência de Equipa de Casa Marca o Respetivo 1.º Golo

Tabela 4.2: Quartis de Tempo de Sobrevivência para Equipa de Fora Marca Respetivo 1.º Golo

Liga	1.º Quartil	Tempo Mediano	3.º Quartil
Portugal	29	66	ND
Espanha	28	62.5	ND
Alemanha	26	52	ND
Inglaterra	28	61	ND
França	31	65	ND
Itália	28	61	ND

ND - não disponível

## 4.2 1.º Golo do Jogo

Nesta secção apresentaremos os resultados referentes ao 1.º golo do jogo. Assim começamos com a estimativa de Kaplan-Meier da função de sobrevivência de dois tipos de eventos: 1.º golo do jogo marcado pela equipa de casa e 1.º golo do jogo marcado pela equipa de fora. Saliente-se que estes dois eventos coincidem com o 1.º golo do jogo ao invés do que sucede com os eventos que analisámos na secção anterior.

### 4.2.1 Estimativa de Kaplan-Meier da Função de Sobrevivência de 1.º Golo Marcado pela Equipa de Casa

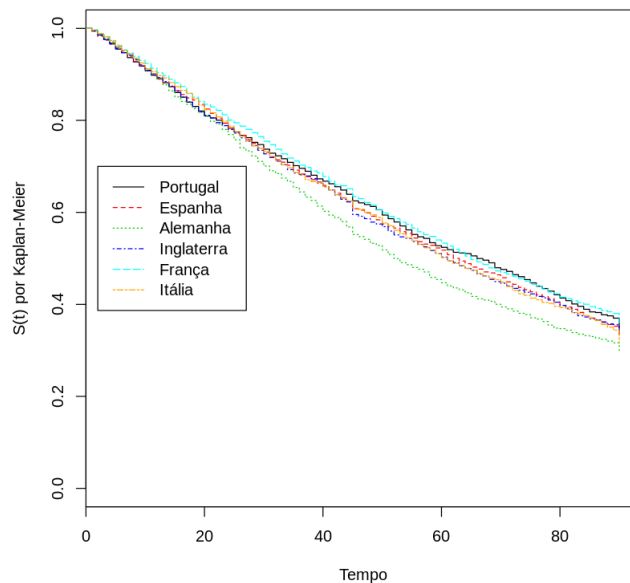


Figura 4.2: Estimativa de Kaplan-Meier da Função de Sobrevivência de Equipa de Fora Marca o Respetivo 1.º Golo

Importa explicitar previamente aqui que, para efeitos deste evento, foram censurados os dados referentes a jogos em que não ocorreu qualquer golo bem como os referentes a jogos em que o 1.º golo do jogo não foi marcado pela equipa de casa. Nos casos de jogos em que não se registou qualquer golo a observação é censurada correspondendo ao tempo máximo (90 minutos) e nos casos em que o evento de interesse foi precedido por outro evento (e.g. para o caso de 1.º golo marcado pela equipa de casa, a equipa de fora ter marcado um golo) considerou-se que “a observação do indivíduo terminou antes da ocorrência do acontecimento”. Os dados foram assim objeto de censura à direita, à semelhança do procedimento descrito na secção 4.1. Foi ainda criada uma variável “ligas” que permitiu identificar a que liga pertenciam os jogos em análise.

#### 4.2.1.1 Quartis de Tempo de Sobrevivência

Os quartis para este evento foram os indicados na tabela 4.3.

#### 4.2.1.2 Função de Sobrevivência

Considerando estes dados procedeu-se ao cálculo da estimativa de Kaplan-Meier e à aplicação do Teste de Log-Rank para comparação das funções sobrevivência das diversas ligas. Neste caso o resultado obtido foi correspondeu um valor-p de 0.0793 que indica que não existem diferenças significativas entre as diferentes ligas. O gráfico da função de sobrevivência obtida por via do Estimador de Kaplan-Meier foi o apresentado na figura 4.3.

Tabela 4.3: Quartis de Tempo de Sobrevivência para a Estimativa de Kaplan-Meier de 1.º Golo Marcado pela Equipa de Casa

Liga	1.º Quartil	Tempo Mediano	3.º Quartil
Portugal	22	47	ND
Espanha	20	45	90
Alemanha	19	43	83
Inglaterra	22	46	86
França	23	50	90
Itália	21	47	88

ND - não disponível

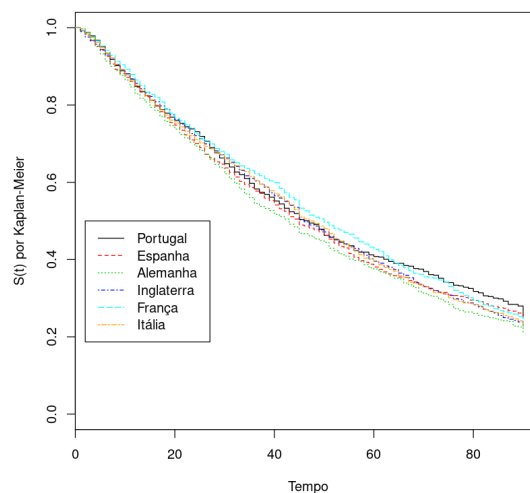


Figura 4.3: Estimativa de Kaplan-Meier da Função de Sobrevivência de 1.º Golo Marcado por Equipa de Casa

## 4.2.2 Estimativa de Kaplan-Meier da Função de Sobrevivência de 1.º Golo Marcado pela Equipa de Fora

Neste caso o evento de interesse correspondeu a um registo do tempo de jogo em que ocorreu o 1.º golo marcado pela equipa de “fora” sendo este registo observado, mas também foram registados os tempos em que ocorreu um 1.º golo marcado pela equipa de “casa” sendo este registo marcado como censurado. Foram também censurados os registos relativos a jogos em que não ocorreu qualquer golo.

A exemplo do evento referido na subsecção anterior foi ainda criada uma variável “ligas” que permitiu identificar a que liga pertenciam os jogos.

### 4.2.2.1 Quartis de Tempo de Sobrevivência

Os quartis para esta evento foram os indicados na tabela 4.4.

Tabela 4.4: Quartis de Tempo de Sobrevivência para a Estimativa de Kaplan-Meier de 1.º Golo Marcado pela Equipa de Fora

Liga	1.º Quartil	Tempo Mediano	3.º Quartil
Portugal	28	61	ND
Espanha	28	61	ND
Alemanha	26	52	ND
Inglaterra	28	63	ND
França	31	64	ND
Itália	28	64	ND

ND - não disponível

### 4.2.2.2 Função de Sobrevivência

Considerando estes dados procedeu-se ao ajustamento do Estimador de Kaplan-Meier e à aplicação do Teste de Log-Rank. O gráfico da função de sobrevivência obtida pelo estimador de Kaplan-Meier foi o que se apresenta na figura 4.4. Neste caso obtivemos um valor-p de 0.038 o que indica que, ao invés do 1.º golo marcado por equipa de casa analisado anteriormente, existem diferenças significativas para os jogos das diferentes ligas.



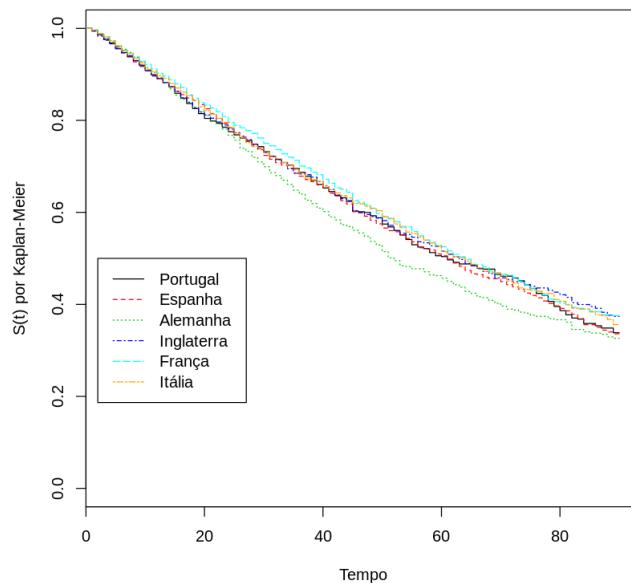


Figura 4.4: Estimativa de Kaplan-Meier de 1.º Golo Marcado por Equipa de Fora

### 4.3 Modelo de Cox para o 1.º Golo do Jogo

Depois de obter a estimativa de Kaplan-Meier para a função de sobrevivência relativa aos eventos do 1.º golo do jogo (equipa de casa ou equipa de fora) procedemos ao ajustamento dos dados ao modelo de riscos proporcionais de Cox.

Nesta secção vamos então apresentar os resultados referentes à aplicação desta metodologia, para esse efeito iniciamos com a verificação da hipótese de riscos proporcionais aplicada às diversas variáveis tal como estão listadas na tabela 4.5.

Tabela 4.5: Listagem de Variáveis

Coluna	Variável	Descrição/Observações
1	<i>tempo</i>	Tempo até ao evento em minutos
2	<i>censura</i>	0=Censurada, 1=Estudo
3	<i>equipa</i>	0=casa, 1=fora
4	<i>odds</i>	Vantagem da equipa de casa atribuída pelas casas de apostas
5	<i>ligas</i>	Variável categórica correspondente às diversas ligas profissionais de membros da U.E.: 1=Portugal, 2=Espanha, 3=Alemanha, 4=Inglaterra, 5=França, 6=Itália

Após a verificação da hipótese de riscos proporcionais procedemos à construção de modelos que incluam exclusivamente covariáveis que respeitam aquela hipótese.

Na construção destes modelos passamos por um processo de seleção de variáveis e

em alguns casos apresentamos a expressão da função de sobrevivência em função da afetação de diversos valores às covariáveis.

### 4.3.1 Hipótese de Riscos Proporcionais

Para a análise dos resíduos de Schoenfeld produziram-se os gráficos que se apresentam na figura 4.5.

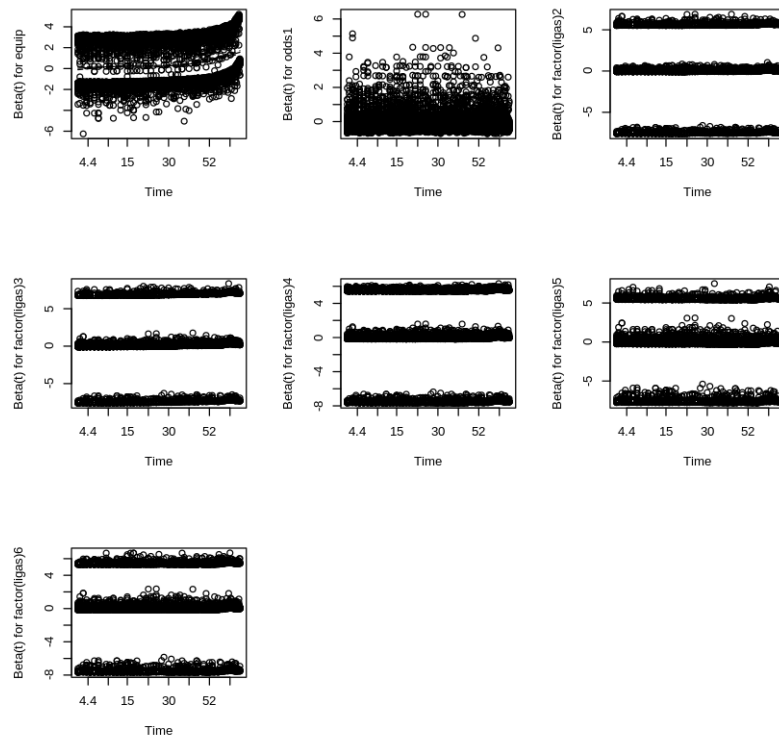


Figura 4.5: Resíduos de Schoenfeld para Modelo de Cox do 1.º Golo do Jogo

Da observação da maioria dos gráficos pode-se concluir que a curva adaptada para os resíduos de Schoenfeld é praticamente constante, i.e., estamos na presença de uma linha horizontal, nas variáveis “ligas” e “odds”. Por outro lado, a variável equipa ("equip") não apresenta um gráfico de denote uma constante através de uma linha horizontal.

O que nos levou a concluir que o mais adequado seria ajustar dois modelos diferentes, sendo um para golo marcado pela equipa de casa e outro para golo marcado por equipa de fora.

Por esse motivo foram construímos dois modelos de riscos proporcionais de Cox para dois tipos diferentes de eventos de golo em função da equipa que marcou e conforme descrito no parágrafo anterior.

### 4.3.2 Equipa de Casa

Nas secções seguintes vamos apresentar os resultados referentes ao evento 1.º golo do jogo marcado pela equipa de casa. Iniciamos com o processo de seleção de variáveis e terminaremos com as funções taxa de falha e taxa de falha acumulada.

#### 4.3.2.1 Selecção de Variáveis

Para este modelo foram inicialmente obtidos os resultados correspondentes aos coeficientes e valores-p indicados na tabela 4.6.

Tabela 4.6: Modelo de Cox para 1.º Golo - Casa: modelo inicial

Covariável	$\hat{\beta}_1$	valor-p
odds	-0.2037	$< 2 \times 10^{-16}$
ligas=Espanha	0.0170	0.7532
ligas=Alemanha	0.1082	0.0382
ligas=Inglaterra	0.0214	0.6663
ligas=França	-0.0523	0.2950
ligas=Itália	0.0146	0.7681

O processo de seleção de covariáveis seguiu os passos indicados na metodologia e no caso do modelo casa estão sumariados os passos na tabela 4.7.

No primeiro passo retirámos a covariável com maior valor-p e procedemos ao cálculo da teste de razão de verosimilhanças, assim retirámos  $ligas = Itália$  e obtivemos um TRV para o valor de 0.087 que comparada com o valor de  $\chi_2^2 = 0.3781$  é manifestamente inferior o que quer dizer que não existe significância estatística para afirmar que o novo modelo é diferente do modelo inicial e prosseguimos com este método até obter um TRV com valor superior ao de  $\chi_p^2$  com  $p$  sendo o número de covariáveis do modelo.

Tabela 4.7: Modelo de Cox para 1.º Golo - Casa: Passos de Eliminação de Covariáveis

Passo	Covariável retirada	TRV	$\chi_p^2$	Decisão
1.º	$ligas = Itália$	0.0871	0.670	Continuar
2.º	$ligas = Espanha$	0.1294	0.670	Continuar
3.º	$ligas = Inglaterra$	0.1986	0.670	Continuar
4.º	$ligas = França$	3.564	0.670	<b>Parar</b>

Na sequência deste processo o modelo obtido encontra-se resumido na tabela 4.8.

Tabela 4.8: Modelo de Cox para 1.º Golo - Casa: modelo final

Covariável	$\hat{\beta}_1$	valor-p
odds	-0.2037	$< 2 \times 10^{-16}$
$ligas = Alemanha$	0.094	0.0174
$ligas = França$	-0.0665	0.0683

Para o modelo casa, a figura 4.6 apresenta as funções correspondentes às 3 ligas do modelo casa com o valor da variável *odds* fixado em 1 (para permitir uma melhor comparação) e na figura 4.7 as funções referentes a vários valores de *odds*.

No gráfico da função de sobrevivência referente a Portugal deve entender-se como correspondente também às ligas de Espanha, Inglaterra e Itália.

Deste gráfico conclui-se que nos jogos das ligas de França e Alemanha as probabilidades de sobrevivência a 1.º golo marcado por equipa de casa são opostamente diferentes com maiores probabilidades em França e menores na Alemanha, tendo as restantes ligas um valor intermédio.

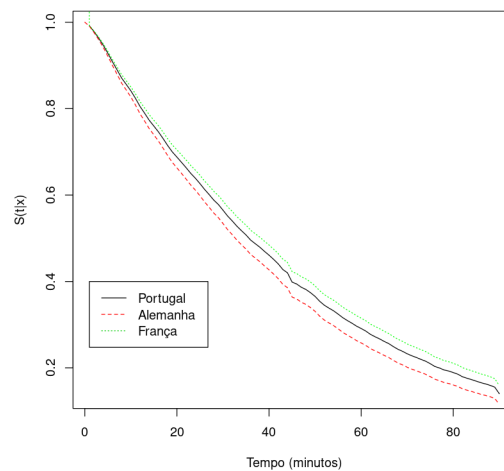


Figura 4.6: Função de Sobrevivência pelo Modelo de Cox para 1.ºGolo-Casa: Covariável ligas, odds fixa.

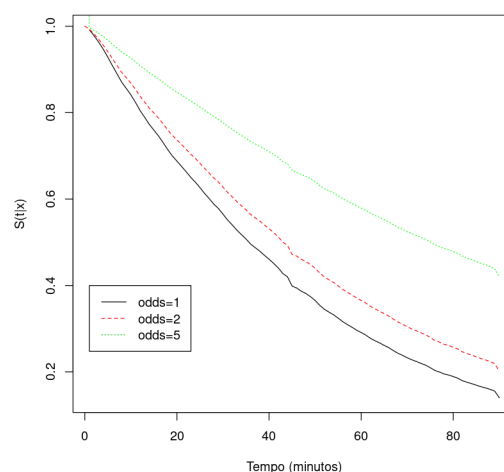


Figura 4.7: Função de Sobrevivência pelo Modelo de Cox para 1.ºGolo-Casa: Covariável odds, liga fixa

Da observação do gráfico da figura 4.7 verifica-se que o maior valor de *odds* aumenta

a probabilidade de sobrevivência ao evento 1.º golo marcado por equipa de casa, o que faz sentido pois as *odds* estão relacionadas com expectativa das casas de apostas relativamente à vitória da equipa de casa.

#### 4.3.2.2 Funções Taxa de Falha e Taxa de Falha Acumulada

Para terminar esta subsecção apresentamos a função taxa de falha para o evento 1.º golo marcado pela equipa de casa que se encontra na figura 4.8.

Relembramos que a taxa de falha corresponde à variação instantânea de probabilidade de ocorrência do evento de interesse, neste caso por minuto.

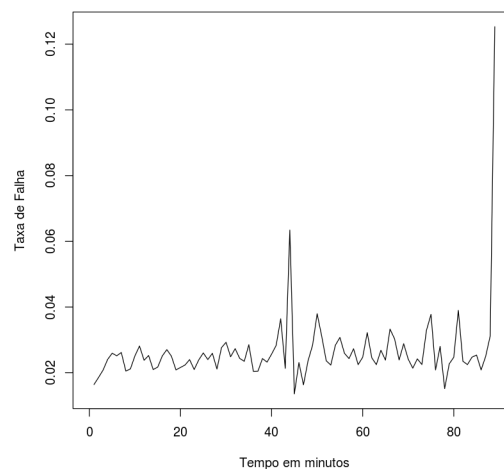


Figura 4.8: Função Taxa de Falha basal para 1.ºGolo- Equipa de Casa

Saliente-se que os valores elevados obtidos nos minutos 45 e 90 se devem na realidade ao facto de nestes estarem também incluídos os golos marcados nos denominados períodos de desconto, estes definidos pelo árbitro do jogo para compensar interrupções durante, respetivamente, a 1.ª e 2.ª partes do jogo.

A função taxa de falha acumulada para o mesmo evento (1.º golo - casa) encontra-se na figura 4.9.

#### 4.3.3 Equipa de Fora

Nas secções seguintes vamos apresentar os resultados referentes ao evento 1.º golo do jogo marcado pela equipa de fora. Iniciamos com o processo de seleção de variáveis e terminaremos com as funções taxa de falha e taxa de falha acumulada.

##### 4.3.3.1 Seleção de Variáveis

Para este modelo foram inicialmente obtidos os resultados correspondentes aos coeficientes e valores-p indicados na tabela 4.9.

Tal como para o modelo casa seguiram-se os passos indicados na metodologia e que neste caso estão sumariados na tabela 4.10.

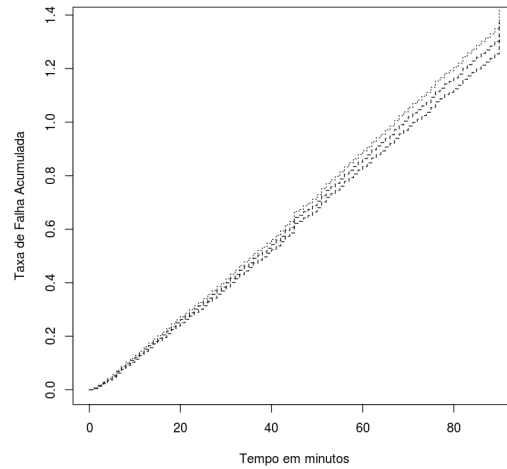


Figura 4.9: Função Taxa de Falha Acumulada para 1.º Golo-Equipa de Casa

Tabela 4.9: Modelo de Cox para 1.º Golo - Fora: modelo inicial

Covariável	$\hat{\beta}_1$	valor-p
<i>odds</i>	0.1037	$< 2 \times 10^{-16}$
<i>ligas = Espanha</i>	0.0337	0.5580
<i>ligas = Alemanha</i>	0.1181	0.0503
<i>ligas = Inglaterra</i>	0.0083	0.8854
<i>ligas = França</i>	0.0234	0.6870
<i>ligas = Itália</i>	0.0463	0.4213

Tabela 4.10: Modelo de Cox para 1.º Golo - Fora: Passos de Seleção de Covariáveis

Passo	Covariável retirada	TRV	$\chi_p^2$	Decisão
1.º	<i>ligas = Inglaterra</i>	0.0208	0.670	Continuar
2.º	<i>ligas = França</i>	0.1725	0.670	Continuar
3.º	<i>ligas = Espanha</i>	0.4270	0.670	Continuar
4.º	<i>ligas = Itália</i>	0.8899	0.670	<b>Parar</b>

Uma vez terminado este processo os coeficientes e valores-p das covariáveis do modelo obtido encontram-se resumidos na tabela 4.11.

Tabela 4.11: Modelo de Cox para 1.º Golo - Fora: modelo final

Covariável	$\hat{\beta}_1$	valor-p
<i>odds</i>	0.1034	$< 2 \times 10^{-16}$
<i>ligas = Alemanha</i>	0.1008	0.0297
<i>ligas = Itália</i>	0.029	0.4951

Com os modelos definidos foi-nos possível obter um gráfico das funções de sobrevivência para diversos valores das covariáveis.

Para o modelo fora, a figura 4.10 contém as funções correspondentes a 4 ligas do modelo fora com valor da variável *odds* fixo em 1.

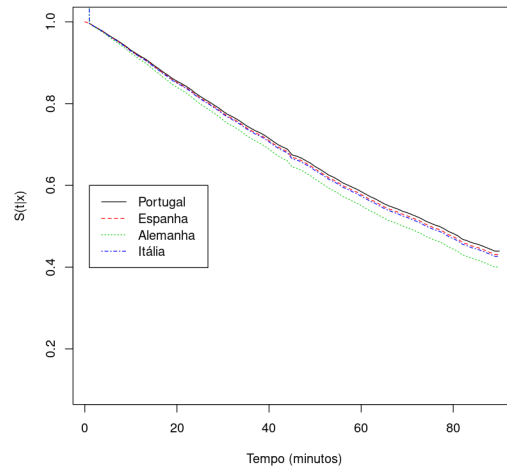


Figura 4.10: Função de Sobrevivência pelo Modelo de Cox para 1.ºGolo - Fora:  
Covariável ligas, *odds* fixa

Também para o mesmo modelo a figura 4.11 apresenta as funções referentes a vários valores de *odds*, deste feita com liga fixa.

Neste caso o gráfico da função de sobrevivência da liga de Portugal também corresponde às outras ligas não indicadas: França e Inglaterra.

Pelo gráfico pode verificar-se que a maioria das ligas têm valor muito aproximado e a liga da Alemanha destaca-se embora ligeiramente.

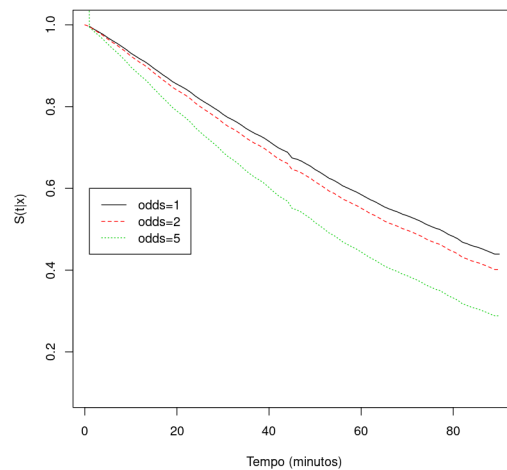


Figura 4.11: Função de Sobrevivência pelo Modelo de Cox para 1.ºGolo - Fora:  
Covariável *odds*, liga fixa

### 4.3.3.2 Funções Taxa de Falha e Taxa de Falha Acumulada

Nesta parte da subsecção apresentamos a função taxa de falha para o evento 1.º golo marcado pela equipa de fora e que se encontra na figura 4.12.

À semelhança do evento descrito na secção anterior, também aqui os valores obtidos para os minutos 45 e 90 correspondem a registos de golos marcados em período de compensação definidos pelo árbitro do jogo.

À semelhança da taxa de falha basal para 1.º golo da equipa de casa não é possível identificar uma nítida tendência de crescimento ao longo do tempo.

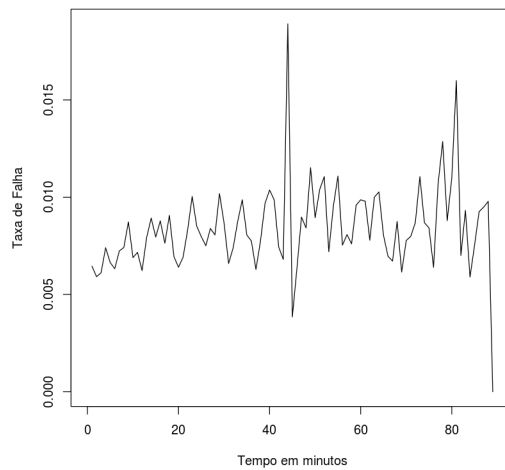


Figura 4.12: Função Taxa de Falha basal para 1.ºGolo - Equipa de Fora

Quanto à função taxa de falha acumulada para o mesmo evento pode-se ver a sua forma na figura 4.13.

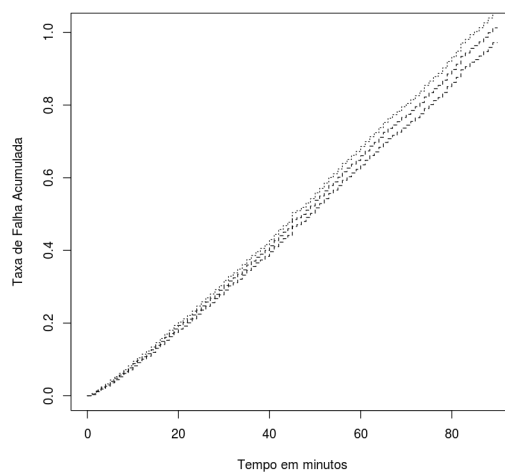


Figura 4.13: Função Taxa de Falha Acumulada para 1.ºGolo - Fora



### 4.3.4 Comparação das Funções Taxa de Falha: Equipa de Casa vs. Equipa de Fora

Nesta parte apresentamos o gráfico que compara as duas funções taxas de falha dos dois eventos: 1.º golo por equipa de casa e 1.º golo por equipa de fora, que se pode ver na figura 4.14.

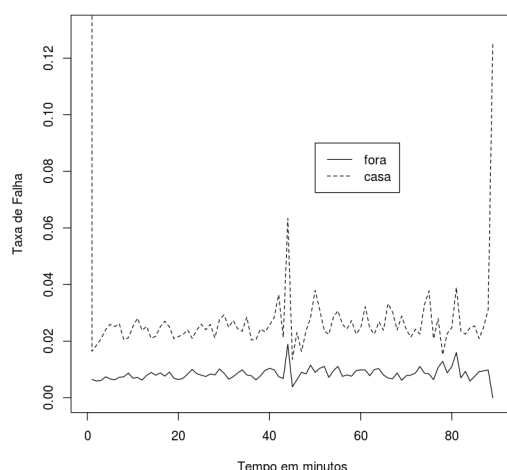


Figura 4.14: Comparação de Funções Taxa de Falha basal para 1.º Golo

Pode-se verificar que o evento referente ao golo da equipa de casa tem uma taxa de falha com valores sempre superiores ao evento referente ao golo da equipa de fora, o que comprova a preponderância atribuída por diversos autores ao fator casa (*Nevo e Ritov, 2012*).

## 4.4 Estimativa de Kaplan-Meier para o 2.º Golo

Nesta secção apresentamos os resultados referentes ao 2.º golo do jogo. Iniciamos com a estimativa de Kaplan-Meier da função de sobrevivência identificando a variável equipa (casa/fora).

Em seguida apresentamos a estimativa de Kaplan-Meier para dois tipos de eventos: 2.º golo do jogo marcado pela equipa de casa e 2.º golo do jogo marcado pela equipa de fora.

### 4.4.1 2.º Golo do Jogo

Para a análise deste evento iniciamos com a apresentação dos quartis do tempo de sobrevivência e posteriormente apresentamos a estimativa da função de sobrevivência.

Assim, os quartis do tempo de sobrevivência para este evento foram os indicados na tabela 4.12.

A figura 4.15 apresenta a estimativa de Kaplan-Meier da função de sobrevivência para este evento.

Tabela 4.12: Quartis do Tempo de Sobrevivência do 2.º Golo do Jogo

Liga	1.º Quartil	Tempo Mediano	3.º Quartil
Portugal	39	61	90
Espanha	36	56	85
Alemanha	34	53	79
Inglaterra	37	56	87
França	40	61	90
Itália	36	58	86

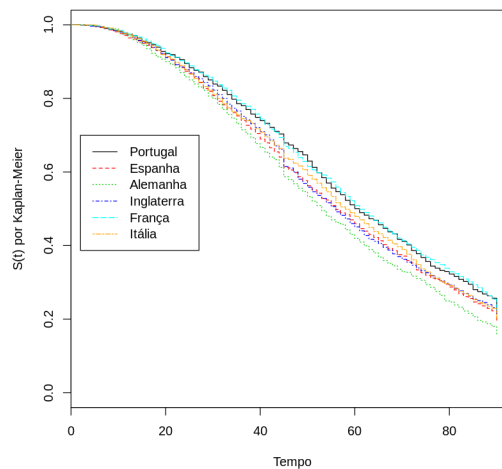


Figura 4.15: Estimativa de Kaplan-Meier da Função de Sobrevivência para 2.º Golo do Jogo

Saliente-se que para efeitos das análises referentes ao 2.º golo são considerados os jogos em que ocorreu pelo menos um golo, os únicos que se pode dizer terem atingido um estado que reunia condições para a ocorrência deste evento.

#### 4.4.2 Equipa de Casa

De seguida apresentamos a análise dos resultados referentes a este evento salientando previamente que para este efeito foram censurados os jogos em que ocorreu apenas um golo e os jogos em que a equipa de fora marcou o 2.º golo.

Os quartis do tempo de sobrevivência foram os indicados na tabela 4.13.

##### 4.4.2.1 Função de Sobrevivência

Em seguida apresentamos os resultados do cálculo da estimativa de Kaplan-Meier cuja expressão se pode encontrar na figura 4.16.

Tabela 4.13: Quartis do Tempo de Sobrevivência do 2.º Golo do Jogo Marcado pela Equipa de Casa

Liga	1.º Quartil	Tempo Mediano	3.º Quartil
Portugal	52	85	90
Espanha	45	79	90
Alemanha	45	78	90
Inglaterra	48	82	90
França	51	85	90
Itália	51	81	90

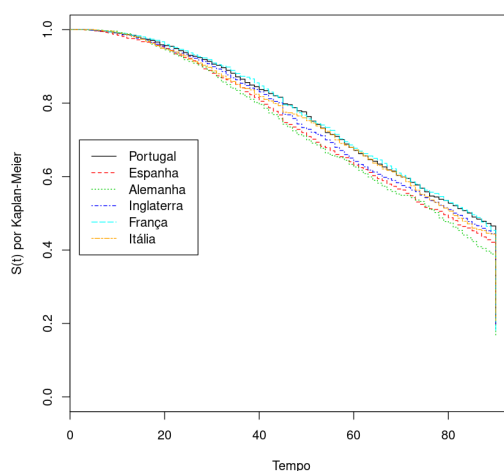


Figura 4.16: Estimativa de Kaplan-Meier da Função de Sobrevivência para 2.º Golo do Jogo Marcado pela Equipa de Casa

### 4.4.3 Equipa de Fora

Segue-se a análise dos resultados referentes a este evento (2.º golo do jogo marcado pela equipa de fora) também destacando que para este efeito foram censurados os jogos em que ocorreu apenas um golo e os jogos em que a equipa de casa marcou o 2.º golo.

Os quartis do tempo de sobrevivência foram os indicados na tabela 4.14.

Tabela 4.14: Quartis do Tempo de Sobrevivência do 2.º Golo do Jogo Marcado pela Equipa de Fora

Liga	1.º Quartil	Tempo Mediano	3.º Quartil
Portugal	60	ND	ND
Espanha	57	ND	ND
Alemanha	52	85	ND
Inglaterra	53	ND	ND
França	61	ND	ND
Itália	55	90	ND

#### 4.4.3.1 Função de Sobrevivência

Em seguida apresentamos os resultados do cálculo da estimativa de Kaplan-Meier que estão expressos na figura 4.17.

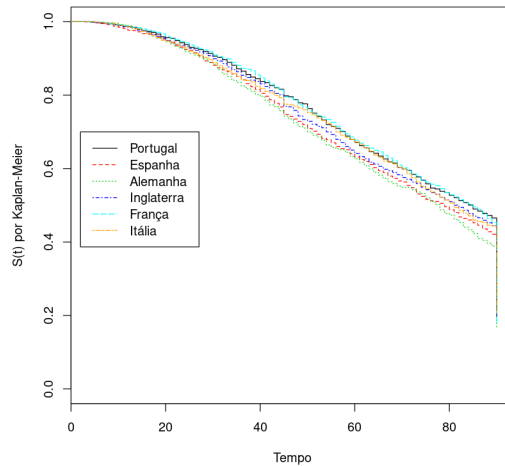


Figura 4.17: Estimativa de Kaplan-Meier da Função de Sobrevivência para 2.º Golo do Jogo Marcado pela Equipa de Fora

## 4.5 Modelo de Cox para o 2.º Golo

Nesta secção vamos apresentar os resultados obtidos para o ajustamento do modelo de Cox aos dados referentes ao 2.º golo do jogo.

Iniciamos com a verificação da hipótese de riscos proporcionais, prosseguimos com o processo de seleção de variáveis e completamos com a apresentação das funções taxa de falha basal e taxa de falha acumulada.

Uma vez que em secção posterior apresentaremos uma análise dos resultados referentes ao mesmo evento, mas com tempo inicial diferente, salientamos que neste caso se consideraram os eventos com tempos medidos desde o início do jogo, isto é,  $t_0 = 0$ .

### 4.5.1 Hipótese de Riscos Proporcionais

Na figura 4.18 podemos observar os gráficos referentes ao modelo de Cox para o 2.º golo do jogo em que se incluem as seguintes variáveis: equipa (casa/fora), ligas e *odds*.

Em face destes resultados concluímos que as variáveis ligas e *odds* respeitam a hipótese de riscos proporcionais e podem ser incluídas na construção do modelo de Cox, ao invés do que se passa com a variável equipa.

Em face deste resultado e considerando a equipa que marcou o 2.º golo podemos identificar diversos tipos de eventos que passamos a descrever:

1. equipa de casa lidera o marcador (1-0) e equipa de casa marca o 2.º golo (2-0)
2. equipa de casa lidera o marcador (1-0) e equipa de fora marca o 2.º golo (1-1)

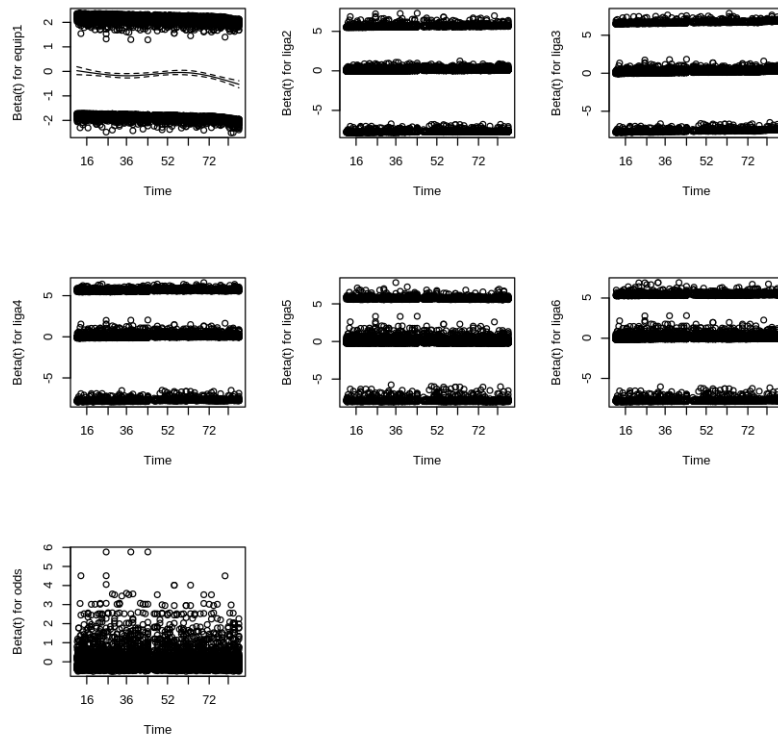


Figura 4.18: Resíduos de Schoenfeld para Modelo de Cox do 2.º Golo do Jogo

3. equipa de fora lidera o marcador (0-1) e equipa de casa marca o 2.º golo (1-1)

4. equipa de fora lidera o marcador (0-1) e equipa de casa marca o 2.º golo (0-2)

então, seguindo esta listagem, vamos passar a construir os diversos modelos para os eventos considerados.

## 4.5.2 Equipa de Casa: de 1-0 para 2-0

De seguida apresentamos resultados referentes ao modelo de Cox para o evento 2.º golo do jogo marcado por equipa de casa.

Salientamos que foram censurados todos os eventos de 2.º golo de jogo correspondentes aos casos: de 1-0 para 1-1, de 0-1 para 1-1 e de 0-1 para 0-2, também foram censurados os jogos em que ocorreu apenas 1 golo.

Iniciamos com o processo de selecção de variáveis e terminamos com as funções taxa de falha e taxa de falha acumulada.

### 4.5.2.1 Selecção de Variáveis

Na tabela 4.15 indicam-se os valores do modelo inicial para o evento de 1-0 para 2-0.

Tabela 4.15: Modelo de Cox para 2.º Golo do Jogo: de 1-0 para 2-0 - modelo inicial

Covariável	$\beta_i$	valor-p
ligas=Espanha	0.0686	0.1691
ligas=Alemanha	0.1415	0.0072
ligas=Inglaterra	0.0352	0.4831
ligas=França	-0.0266	0.5939
ligas=Itália	-0.0032	0.9494
odds	-0.2031	$<2 \times 10^{-16}$

Na sequência deste modelo inicial procedemos ao seguinte processo de seleção de variáveis, como se indica na tabela 4.16, donde se obteve o modelo final como o indicado na tabela 4.17.

Tabela 4.16: Modelo de Cox para 2.º Golo: de 1-0 para 2-0 - Passos de Seleção de Covariáveis

Passo	Covariável retirada	TRV	$\chi_p^2$	Decisão
1.º	ligas=Itália	0.0040	0.670	Continuar
2.º	ligas=França	0.3700	0.670	Continuar
3.º	ligas=Inglaterra	1.4442	0.670	<b>Parar</b>

Tabela 4.17: Modelo de Cox para 2.º Golo do Jogo: de 1-0 para 2-0 - modelo final

Covariável	$\beta_i$	valor-p
ligas=Espanha	0.0797	0.0359
ligas=Alemanha	0.1525	0.0002
ligas=Inglaterra	0.0462	0.2277
odds	-0.2030	$<2 \times 10^{-16}$

#### 4.5.2.2 Funções Taxa de Falha e Taxa de Falha Acumulada

As funções taxa de falha basal e taxa de falha acumulada correspondem às expressas nas figuras 4.19 e 4.20.

Quanto à função taxa de falha basal verificamos um ligeiro crescimento ao longo do tempo de jogo, o que faz sentido indiciando um ascendente da equipa de casa em relação à progressão do jogo.

O valor do minuto final deve ser ignorado por incluir eventos que correspondem a mais do que 1 minuto de tempo.

#### 4.5.3 Equipa de Casa: de 0-1 para 1-1

Nesta secção vamos apresentar os resultados referentes ao evento 2.º golo do jogo marcado por equipa de casa.

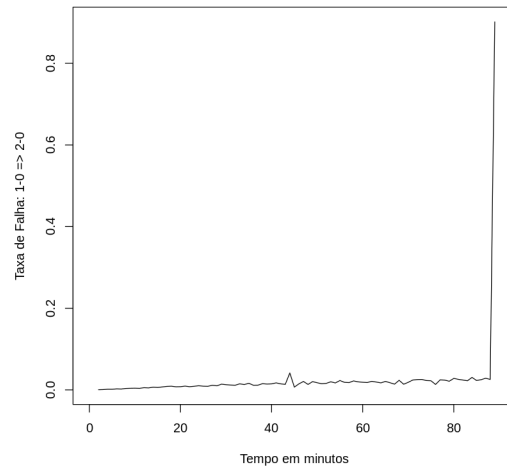


Figura 4.19: Taxa de Falha do 2.º Golo do Jogo: de 1-0 para 2-0

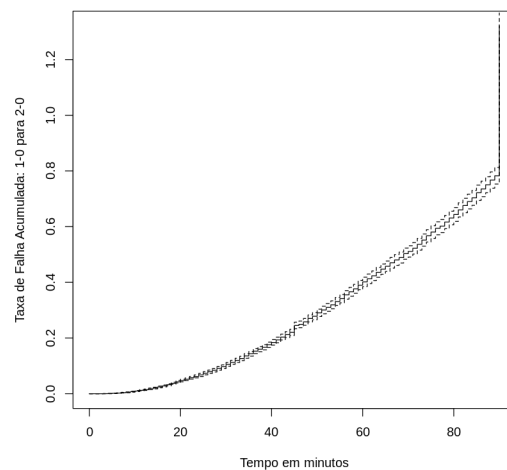


Figura 4.20: Taxa de Falha Acumulada do 2.º Golo do Jogo: de 1-0 para 2-0

Salientamos que foram censurados todos os eventos de 2.º golo de jogo que corresponderam aos casos: de 1-0 para 2-0, de 1-0 para 1-1 e de 0-1 para 0-2, também foram censurados os jogos em que ocorreu apenas 1 golo.

Iniciamos com o processo de seleção de variáveis e continuamos até à taxa de falha e taxa de falha acumulada.

#### 4.5.3.1 Seleção de Variáveis

Para este evento considerou-se o processo de seleção de variáveis partindo de um modelo inicial com os coeficientes indicados na tabela 4.18.

Os passos de seleção de variáveis foram os indicados na tabela 4.19.

O modelo final para este evento foi o indicado na tabela 4.20.

Tabela 4.18: Modelo de Cox para 2.º Golo do Jogo: de 1-0 para 2-0 - modelo inicial

Covariável	$\beta_i$	valor-p
ligas=Espanha	0.0496	0.571
ligas=Alemanha	0.1451	0.111
ligas=Inglaterra	-0.0434	0.626
ligas=França	-0.0423	0.630
ligas=Itália	0.0194	0.824
odds	-0.0122	0.336

Tabela 4.19: Modelo de Cox para 2.º Golo: de 0-1 para 1-1 - Passos de Selecção de Covariáveis

Passo	Covariável retirada	TRV	$\chi_p^2$	Decisão
1.º	ligas=Itália	0.0497	0.670	Continuar
2.º	ligas=França	0.2854	0.670	Continuar
3.º	odds	0.9848	0.670	<b>Parar</b>

Tabela 4.20: Modelo de Cox para 2.º Golo do Jogo: de 0-1 para 1-1 - modelo final

Covariável	$\beta_i$	valor-p
ligas=Alemanha	0.1204	0.0932
ligas=Inglaterra	-0.0681	0.3242
ligas=França	-0.0670	0.3206
odds	-0.0124	0.3279

#### 4.5.3.2 Funções Taxa de Falha e Taxa de Falha Acumulada

A função taxa de falha basal deste evento encontra-se na figura 4.21.

Pode-se verificar uma tendência de crescimento ao longo do tempo, que é comum a todos os eventos correspondentes ao 2.º golo do jogo e que principalmente indicia o esforço para a produção de eventos de golo ao longo do tempo.

A função taxa de falha acumulada encontra-se na figura 4.22.

#### 4.5.4 Equipa de Fora: de 1-0 para 1-1

Nesta secção vamos apresentar os resultados referentes ao evento 2.º golo do jogo, neste caso de 1-0 para 1-1, sendo que se ocorrer o 2-0 este foi considerado censurado.

Salientamos que foram censurados todos os eventos de 2.º golo de jogo correspondentes aos casos: de 1-0 para 2-0, de 0-1 para 1-1 e de 0-1 para 0-2, também foram censurados os jogos em que ocorreu apenas 1 golo.

Iniciamos com o processo de seleção de variáveis e completamos com a apresentação das funções taxa de falha e taxa de falha acumulada.



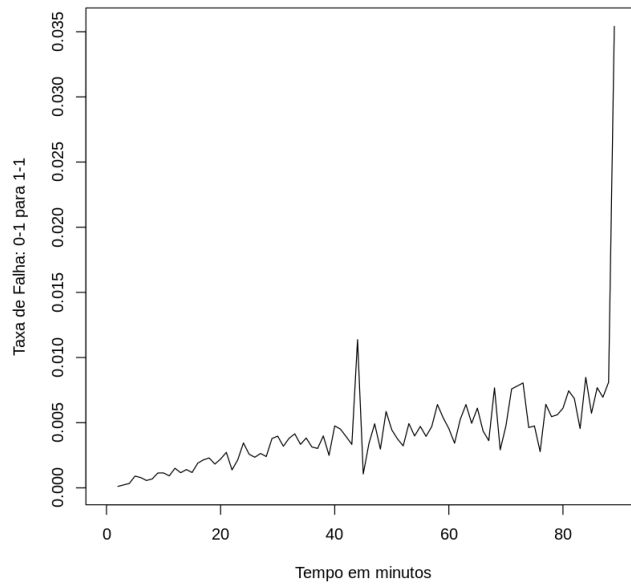


Figura 4.21: Função Taxa de Falha de 2.º Golo do Jogo: de 0-1 para 1-1

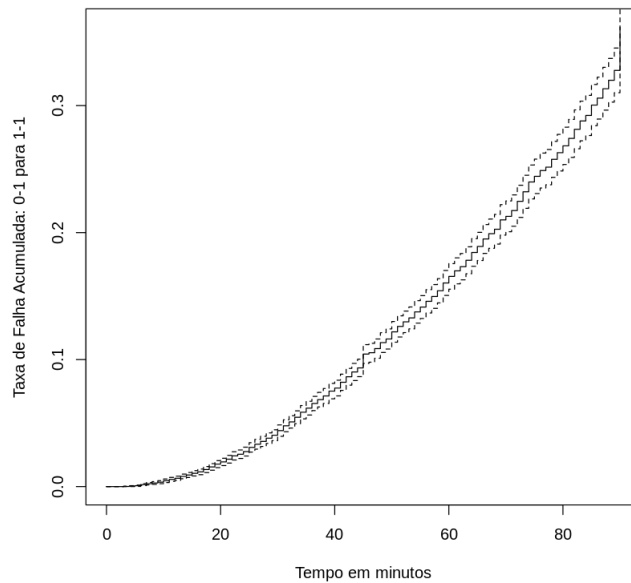


Figura 4.22: Função Taxa de Falha de 2.º Golo do Jogo: de 0-1 para 1-1

#### 4.5.4.1 Seleção de Variáveis

O modelo inicial para este evento foi o indicado na tabela 4.21.

Os passos de seleção de variáveis foram os indicados na tabela 4.22.

Após o processo de seleção de variáveis o modelo final obtido foi o indicado na

Tabela 4.21: Modelo de Cox para 2.º Golo do Jogo: de 1-0 para 1-1 - modelo inicial

Covariável	$\beta_i$	valor-p
ligas=Espanha	0.0105	0.9058
ligas=Alemanha	0.2548	0.0043
ligas=Inglaterra	0.1185	0.1700
ligas=França	-0.0514	0.5608
ligas=Itália	0.1765	0.0372
odds	-0.0142	0.2577

Tabela 4.22: Modelo de Cox para 2.º Golo: de 1-0 para 1-1 - Passos de Seleção de Covariáveis

Passo	Covariável retirada	TRV	$\chi_p^2$	Decisão
1.º	ligas=Espanha	0.0140	0.670	Continuar
2.º	ligas=França	0.6138	0.670	Continuar
3.º	odds	1.219	0.670	Parar

tabela 4.23.

Tabela 4.23: Modelo de Cox para 2.º Golo do Jogo: de 1-0 para 1-1 - modelo final

Covariável	$\beta_i$	valor-p
ligas=Alemanha	0.27	0.00001
ligas=Inglaterra	0.1338	0.0416
ligas=Itália	0.1918	0.0025
odds	-0.0136	0.2769

#### 4.5.4.2 Funções Taxa de Falha e Taxa de Falha Acumulada

A função taxa de falha deste evento encontra-se expressa na figura 4.23.

De forma similar aos eventos descritos anteriormente, verifica-se uma tendência de crescimento ao longo do tempo de jogo.

A função taxa de falha acumulada deste evento pode encontrar-se na figura 4.24.

#### 4.5.5 Equipa de Fora: de 0-1 para 0-2

Nesta secção vamos apresentar os resultados referentes ao evento 2.º golo do jogo marcado por equipa de fora, neste caso de 1-0 para 2-0.

Salientamos que foram censurados todos os eventos de 2.º golo de jogo correspondentes aos casos: de 1-0 para 1-1, de 0-1 para 1-1 e de 1-0 para 2-0, também foram censurados os jogos em que ocorreu apenas 1 golo.

Iniciamos com o processo de seleção de variáveis e finalizamos com a apresentação das funções taxa de falha e taxa de falha acumulada.

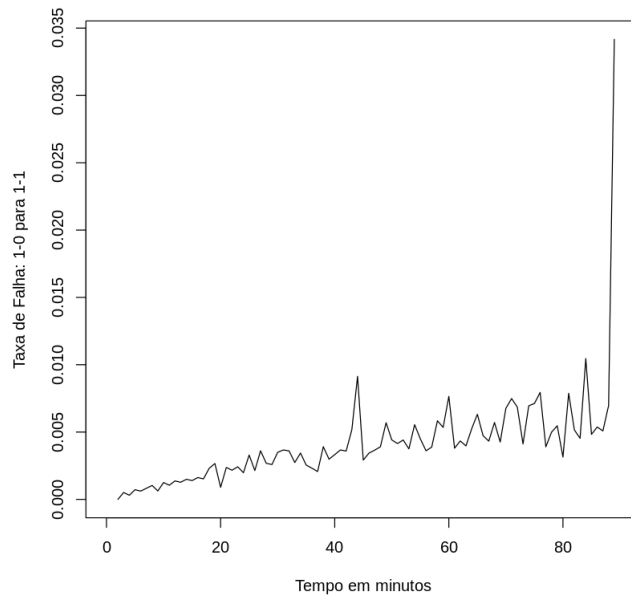


Figura 4.23: Função Taxa de Falha de 2.º Golo do Jogo: de 1-0 para 1-1

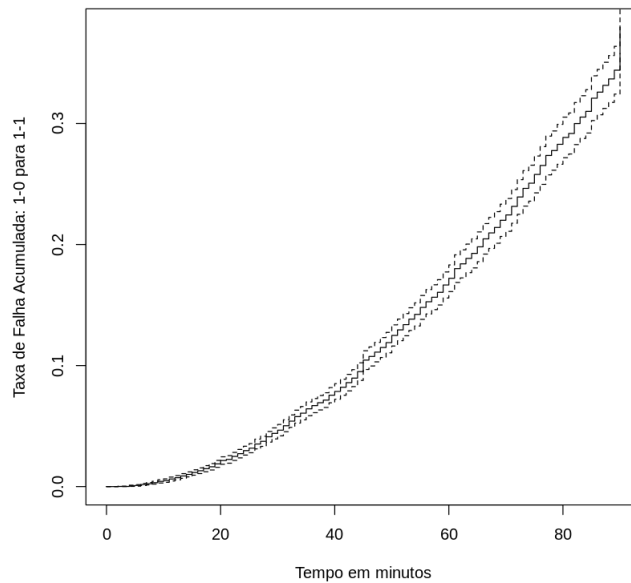


Figura 4.24: Função Taxa de Falha Acumulada de 2.º Golo do Jogo: de 1-0 para 1-1

#### 4.5.5.1 Seleção de Variáveis

O modelo inicial obtido para este evento foi o indicado na tabela 4.24.

O único passo de selecção de covariáveis foi o indicado na tabela 4.25, donde se pode verificar que neste caso não foram retiradas quaisquer covariáveis.

Tabela 4.24: Modelo de Cox para 2.º Golo do Jogo: de 0-1 para 0-2 - modelo inicial

Covariável	$\beta_i$	valor-p
ligas=Espanha	0.2044	0.0366
ligas=Alemanha	0.3676	0.0003
ligas=Inglaterra	0.3420	0.0005
ligas=França	0.2501	0.0128
ligas=Itália	0.2531	0.0112
odds	0.1727	$< 2 \times 10^{-16}$

Tabela 4.25: Modelo de Cox para 2.º Golo: de 0-1 para 0-2 - Selecção de Covariáveis

Passo	Covariável retirada	TRV	$\chi_p^2$	Decisão
1.º	ligas=Espanha	4.421	0.670	<b>Parar</b>

#### 4.5.5.2 Funções Taxa de Falha e Taxa de Falha Acumulada

A função taxa de falha pode-se observar na figura 4.25.

Nesta função verifica-se uma nítida tendência de crescimento, indiciadora que mesmo ao nível da equipa de fora o esforço de marcar um 2.º golo mantém-se ao longo do tempo.

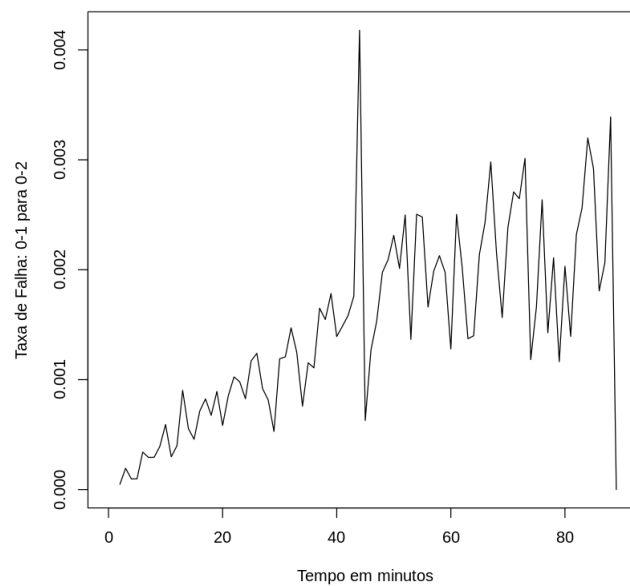


Figura 4.25: Função Taxa de Falha de 2.º Golo do Jogo: de 0-1 para 0-2

A função taxa de falha acumulada pode-se observar na figura 4.26.

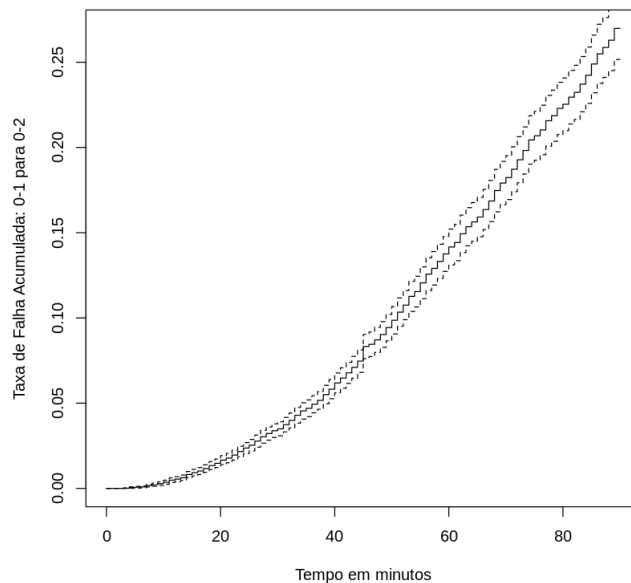


Figura 4.26: Função Taxa de Falha Acumulada de 2.º Golo do Jogo: de 0-1 para 0-2

#### 4.5.6 Comparação de Taxas de Falha: 1-0 para 2-0 vs. 1-0 para 1-1

Nesta parte da secção apresentamos a comparação das funções taxa de falha na figura 4.27 que coloca em oposição dois eventos que podem acontecer após o 1-0 (1.º golo do jogo marcado pela equipa de casa): o 2-0 e o 1-1.

É notório que o fator casa tem uma preponderância relevante e se destaca numa tendência de crescimento mais acentuado na taxa de falha do 1.º destes eventos em relação à taxa de falha do 2.º evento (o 1-0 para 1-1).

#### 4.5.7 Comparação de Taxas de Falha: 0-1 para 0-2 vs. 0-1 para 1-1

Nesta parte da secção apresentamos a comparação das funções taxa de falha na figura 4.27 que coloca em oposição dois eventos que podem acontecer após o 0-1 (1.º golo do jogo marcado pela equipa de fora): o 0-2 e o 1-1.

É notório que o fator casa também aqui revela a sua preponderância com uma tendência de crescimento mais acentuado na taxa de falha do 2.º destes eventos em relação à taxa de falha do 1.º evento (o 0-1 para 0-2).

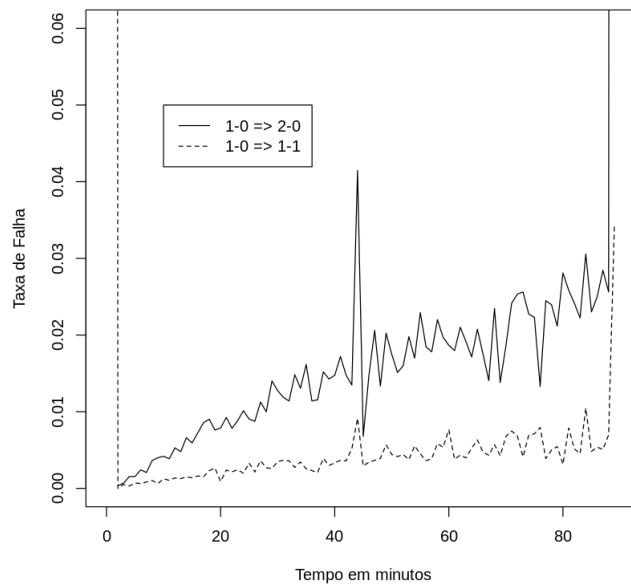


Figura 4.27: Comparação de Taxas de Falha: 1-0 para 2-0 vs. 1-0 para 1-1

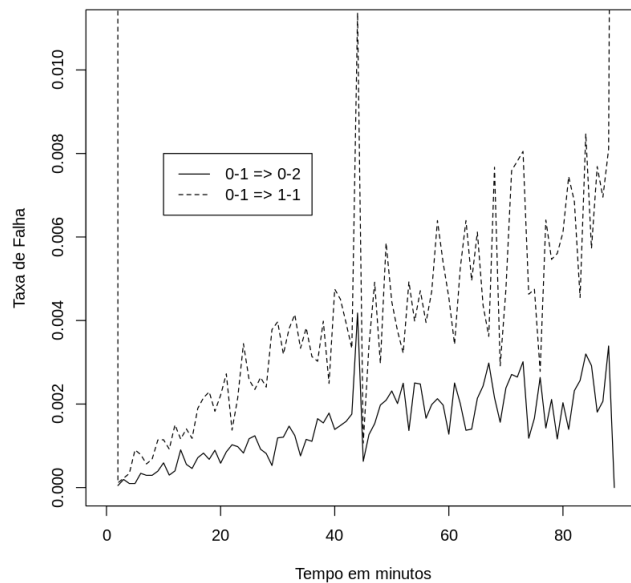


Figura 4.28: Comparação de Taxas de Falha: 0-1 para 0-2 vs. 0-1 para 1-1

## 4.6 Estimativa de Kaplan-Meier da Função de Sobre- vivência de 2.º Golo do Jogo com Tempo Inicial Igual ao Tempo do 1.º Golo

Nesta secção apresentamos os resultados referentes ao 2.º golo do jogo considerando, de forma diferente da secção anterior, como instante inicial o tempo referente ao 1.º

golo do jogo [ $t_0 = t_{golo1}$ ].

À semelhança da secção anterior, consideramos diversos tipos de eventos diferentes, especificamente em função de configurações diferentes na marcha do marcador, nomeadamente:

- o 2.º golo do jogo, qualquer que fosse o resultado anterior,
- o 2.º golo do jogo marcado pela mesma equipa, no caso de esta ser a equipa de casa: de 1-0 para 2-0,
- o 2.º golo do jogo marcado pela mesma equipa, no caso de esta ser a equipa de fora: de 0-1 para 0-2,
- o empate no marcador depois do 1.º golo do jogo marcado pela equipa de casa: do 1-0 para 1-1,
- o empate no marcador depois do 1.º golo do jogo marcado pela equipa de fora: do 0-1 para 1-1.

#### 4.6.1 2.º Golo do Jogo com Tempo Inicial Igual ao Tempo do 1.º Golo

Para análise deste evento iniciamos com a apresentação dos quartis do tempo de sobrevivência e posteriormente apresentamos a estimativa da função de sobrevivência.

Saliente-se que, para efeitos das análises referentes ao 2.º golo, são considerados os jogos em que ocorreu pelo menos um golo, os únicos que se pode dizer terem atingido um estado que reunia condições para a ocorrência deste evento.

Os quartis do tempo de sobrevivência para este tipo de evento foram os indicados na tabela 4.26.

Tabela 4.26: Quartis de Tempo de Sobrevivência de 2.º Golo do Jogo com  $t_0 = t_{golo1}$

Liga	1.º Quartil	Mediana	3.º Quartil
Portugal	13	27	50
Espanha	10	23	45
Alemanha	10	22	41
Inglaterra	11	24	44
França	12	27	49
Itália	11	24	45

##### 4.6.1.1 Função de Sobrevivência

O gráfico da referida estimativa da função de sobrevivência corresponde ao apresentado na figura 4.29.

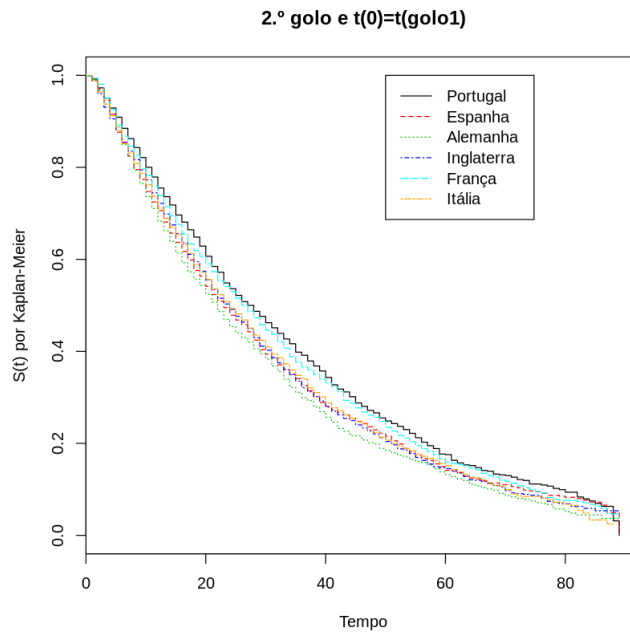


Figura 4.29: Estimativa de Kaplan-Meier da Função de Sobrevivência de 2.º Golo do Jogo com  $t_0=t_{golo1}$

#### 4.6.2 2.º Golo do Jogo Marcado por Equipa de Casa com Tempo Inicial Igual ao Tempo do 1.º Golo

Para a análise deste evento adiantamos inicialmente os quartis do tempo de sobrevivência e prosseguimos para a estimativa da função de sobrevivência.

Os quartis do tempo de sobrevivência para este tipo de evento foram os indicados na tabela 4.26.

Tabela 4.27: Quartis de Tempo de Sobrevivência de 2.º Golo do Jogo Marcado por Equipa de Casa com  $t_0 = t_{golo1}$

Liga	1.º Quartil	Mediana	3.º Quartil
Portugal	18	40	65
Espanha	15	34	62
Alemanha	16	34	61
Inglaterra	17	35	60
França	16	38	61
Itália	16	37	62

##### 4.6.2.1 Função de Sobrevivência

O gráfico da referida estimativa corresponde ao apresentado na figura 4.30.



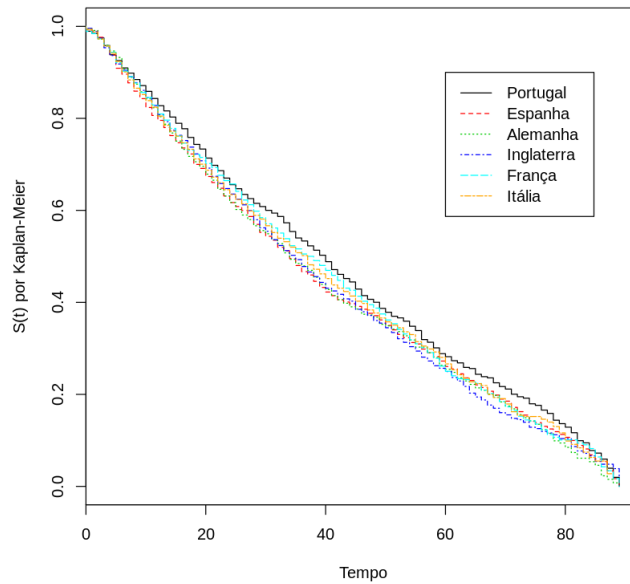


Figura 4.30: Estimativa de Kaplan-Meier da Função de Sobrevivência de 2.º Golo do Jogo com  $t_0=t_{golo1}$

### 4.6.3 2.º Golo do Jogo Marcado por Equipa de Fora com Tempo Inicial Igual ao Tempo do 1.º Golo

À semelhança do referido anteriormente, para a análise deste evento adiantamos inicialmente os quartis do tempo de sobrevivência e prosseguimos para a estimativa da função de sobrevivência.

#### 4.6.3.1 Quartis de Tempo de Sobrevivência

Os quartis do tempo de sobrevivência para este tipo de evento foram os indicados na tabela 4.28.

Tabela 4.28: Quartis de Tempo de Sobrevivência de 2.º Golo do Jogo Marcado por Equipa de Fora com  $t_0 = t_{golo1}$

Liga	1.º Quartil	Mediana	3.º Quartil
Portugal	28	57	ND
Espanha	24	53	ND
Alemanha	21	47	ND
Inglaterra	22	49	ND
França	27	57	ND
Itália	29	49	84

ND - não disponível

### 4.6.3.2 Função de Sobrevivência

O gráfico da referida estimativa corresponde ao apresentado na figura 4.31.

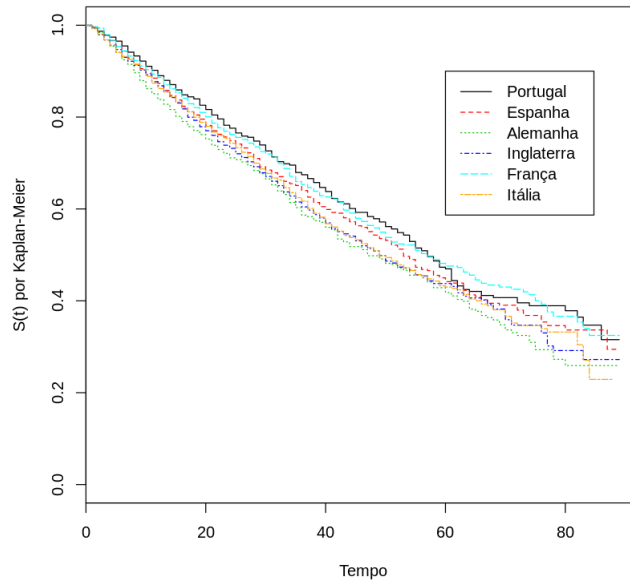


Figura 4.31: Estimativa de Kaplan-Meier da Função de Sobrevivência de 2.º Golo do Jogo com  $t_0=t_{golo1}$

## 4.7 Modelo de Cox para o 2.º Golo do Jogo com Tempo Inicial Igual a Tempo do 1.º Golo

Nesta secção vamos proceder à apresentação dos resultados obtidos do ajustamento do modelo de Cox aos dados referentes ao 2.º golo do jogo, desta feita considerando, por vez do tempo do início do jogo, o momento inicial correspondente ao tempo do 1.º golo do jogo.

Iniciamos com a verificação da hipótese de riscos proporcionais, avançamos depois para o processo de selecção de variáveis e terminamos com a apresentação das funções taxa de falha basal e taxa de falha acumulada.

### 4.7.1 Hipótese de Riscos Proporcionais

Para análise e verificação da hipótese de riscos proporcionais foram produzidos os gráficos da figura 4.32.

Como se pode verificar, todas as variáveis respeitam a hipótese de riscos proporcionais, com exceção da variável equipa.

Neste sentido, à semelhança de secção anterior, foram elaborados modelos para eventos diferentes definidos estes em função da marcha do marcador.

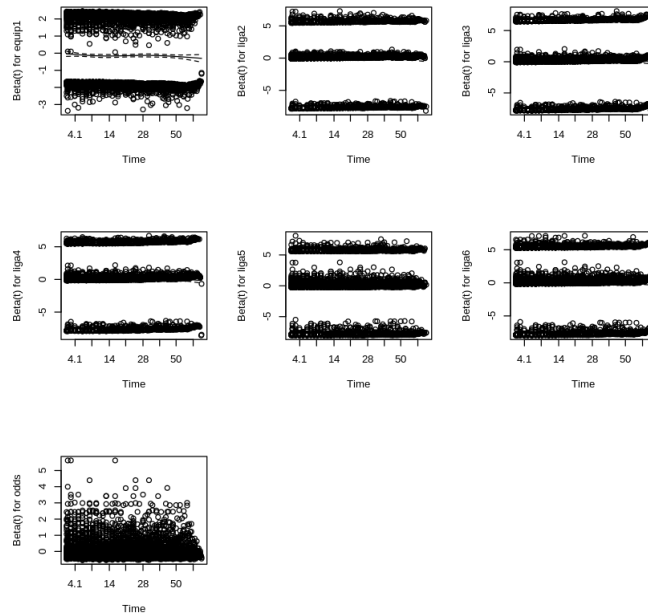


Figura 4.32: Resíduos de Schoenfeld para 2.º golo com  $t_0 = t_{golo1}$

## 4.7.2 Equipa de Casa: de 1-0 para 2-0

Nesta secção vamos apresentar os resultados referentes ao evento 2.º golo do jogo marcado por equipa de casa com  $t_0 = t_{golo1}$ , neste caso do 1-0 para o 2-0.

Salientamos que foram censurados todos os eventos de 2.º golo de jogo correspondentes aos casos: de 1-0 para 1-1, de 0-1 para 1-1 e de 0-1 para 0-2, também foram censurados os jogos em que ocorreu apenas 1 golo.

Iniciamos com o processo de seleção de variáveis e terminamos com a apresentação da taxa de falha e taxa de falha acumulada.

### 4.7.2.1 Seleção de Variáveis

Para efeitos desta análise iniciamos com a inclusão das covariáveis indicadas na tabela 4.29.

Tabela 4.29: Modelo de Cox para 2.º Golo do Jogo de 1-0 para 2-0 com  $t_0 = t_{golo1}$ : modelo inicial

Covariável	$\beta_i$	valor-p
ligas=Espanha	0.0827	0.0976
ligas=Alemanha	0.0926	0.0783
ligas=Inglaterra	0.0945	0.0598
ligas=França	0.0116	0.8167
ligas=Itália	0.0395	0.4294
odds	-0.1935	$<2 \times 10^{-16}$

Tal como realizado para outros eventos já analisados anteriormente, executámos os passos indicados na tabela 4.30.

Tabela 4.30: Modelo de Cox para 2.º Golo do Jogo de 1-0 para 2-0 com  $t_0 = t_{golo1}$ : Passos de Seleção de Covariáveis

Passo	Covariável retirada	TRV	$\chi_p^2$	Decisão
1.º	ligas=França	0.0538	0.670	Continuar
2.º	ligas=Itália	0.6443	0.670	Continuar
3.º	ligas=Inglaterra	3.886	0.670	<b>Parar</b>

Neste sentido o modelo final corresponde ao indicado na tabela 4.31.

Tabela 4.31: Modelo de Cox para 2.º Golo do Jogo de 1-0 para 2-0 com  $t_0 = t_{golo1}$ : modelo final

Covariável	$\beta_i$	valor-p
ligas=Espanha	0.0642	0.0905
ligas=Alemanha	0.0742	0.0735
ligas=Inglaterra	0.0760	0.0473
odds	-0.1936	$<2 \times 10^{-16}$

#### 4.7.2.2 Funções Taxa de Falha e Taxa de Falha Acumulada

Os gráficos correspondentes à função taxa de falha e à função taxa de falha acumulada deste evento estão respetivamente na figura 4.33 e 4.34.

### 4.7.3 Equipa de Casa: de 0-1 para 1-1

Nesta secção vamos apresentar os resultados referentes ao evento 2.º golo do jogo marcado por equipa de casa com  $t_0 = t_{golo1}$ , neste caso do 1-1 para o 1-1.

Salientamos que foram censurados todos os eventos de 2.º golo de jogo correspondentes aos casos: de 1-0 para 2-0, de 1-0 para 1-1 e de 0-1 para 0-2, também foram censurados os jogos em que ocorreu apenas 1 golo.

Iniciamos com o processo de seleção de variáveis e terminamos com a apresentação da taxa de falha e taxa de falha acumulada.

#### 4.7.3.1 Seleção de Variáveis

Para efeitos desta análise iniciamos com o modelo em que se incluem as covariáveis indicadas na tabela 4.32.

Tal como realizado para outros eventos já analisados executámos os passos indicados na tabela 4.33, sendo que neste caso não foi retirada qualquer variável.

Neste sentido o modelo final corresponde ao modelo inicial considerado na tabela 4.18.

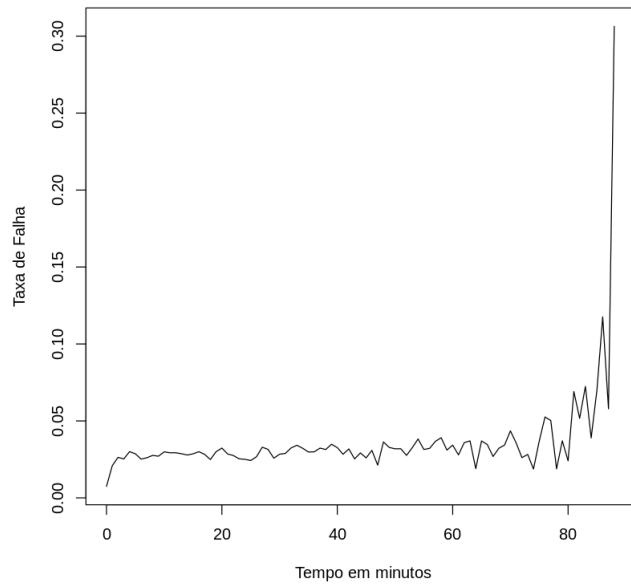


Figura 4.33: Função Taxa de Falha para 2.º Golo de Jogo com  $t_0 = t_{golo1}$

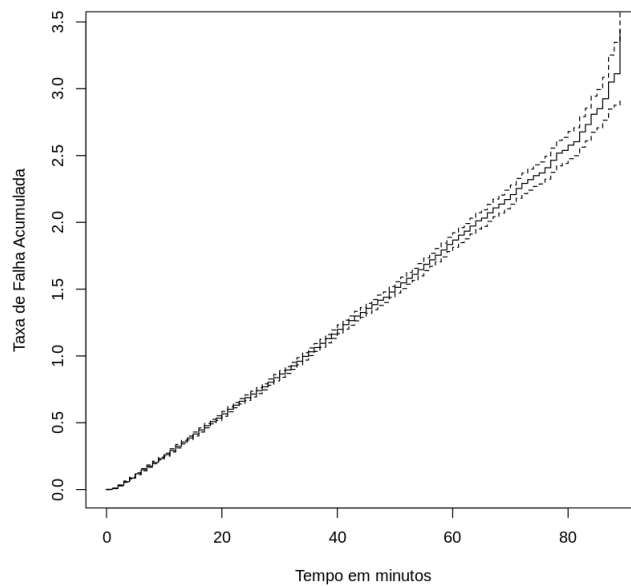


Figura 4.34: Função Taxa de Falha Acumulada para 2.º Golo de Jogo com  $t_0 = t_{golo1}$

#### 4.7.3.2 Funções Taxa de Falha e Taxa de Falha Acumulada

As imagens referentes a estas funções encontram-se na figura 4.35, e também na figura 4.36.

Tabela 4.32: Modelo de Cox para 2.º Golo do Jogo de 0-1 para 1-1 com  $t_0 = t_{golo1}$ : modelo inicial

Covariável	$\beta_i$	valor-p
ligas=Espanha	0.0672	0.442
ligas=Alemanha	0.1221	0.180
ligas=Inglaterra	0.0005	0.996
ligas=França	0.0005	0.995
ligas=Itália	0.0614	0.481
odds	-0.0102	0.414

Tabela 4.33: Modelo de Cox para 2.º Golo do Jogo de 0-1 para 1-1 com  $t_0 = t_{golo1}$ : Passos de Seleção de Covariáveis

Passo	Covariável retirada	TRV	$\chi_p^2$	Decisão
1.º	ligas=Inglaterra	2.8516	0.670	<b>Parar</b>

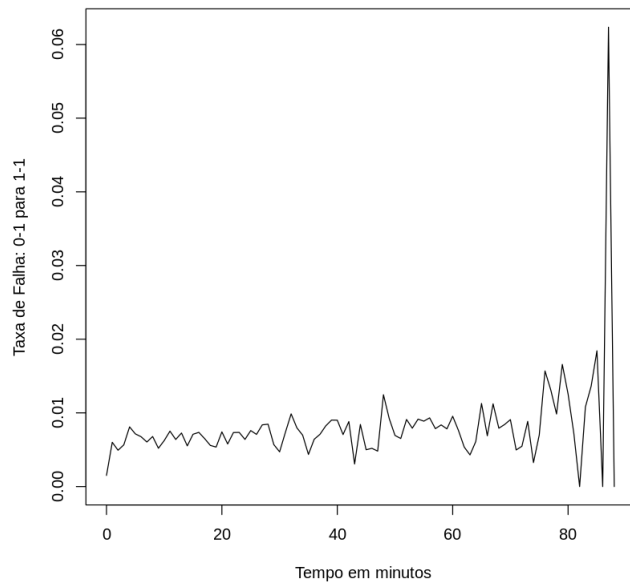


Figura 4.35: Função Taxa de Falha de 2.º Golo: de 0-1 para 1-1 com  $t_0 = t_{golo2}$

#### 4.7.4 Equipa de Casa: de 1-0 para 1-1

Nesta secção vamos apresentar os resultados referentes ao evento 2.º golo do jogo marcado por equipa de fora com  $t_0 = t_{golo1}$ , neste caso do 1-1 para o 1-1.

Salientamos que foram censurados todos os eventos de 2.º golo de jogo correspondentes aos casos: de 1-0 para 2-0, de 0-1 para 1-1 e de 0-1 para 0-2, também foram censurados os jogos em que ocorreu apenas 1 golo.

Iniciamos com o processo de seleção de variáveis e terminamos com a apresentação da taxa de falha e taxa de falha acumulada.

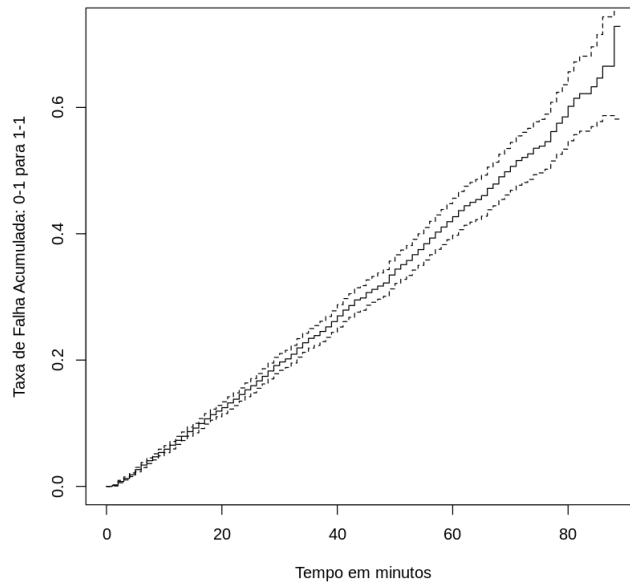


Figura 4.36: Função Taxa de Falha de 2.º Golo: de 0-1 para 1-1 com  $t_0 = t_{golo2}$

#### 4.7.4.1 Seleção de Variáveis

Para efeitos desta análise iniciamos com o modelo em que são incluídas as covariáveis indicadas na tabela 4.34.

Tabela 4.34: Modelo de Cox para 2.º Golo do Jogo de 1-0 para 1-1 com  $t_0 = t_{golo1}$ : modelo inicial

Covariável	$\beta_i$	valor-p
ligas=Espanha	0.0299	0.736
ligas=Alemanha	0.2364	0.008
ligas=Inglaterra	0.1664	0.054
ligas=França	-0.0079	0.929
ligas=Itália	0.2211	0.009
odds	-0.0123	0.320

Tal como realizado para outros eventos já analisados executámos os passos indicados na tabela 4.35.

Tabela 4.35: Modelo de Cox para 2.º Golo do Jogo de 0-1 para 1-1 com  $t_0 = t_{golo1}$ : Passos de Seleção de Covariáveis

Passo	Covariável retirada	TRV	$\chi_p^2$	Decisão
1.º	ligas=França	0.0080	0.670	Continuar
2.º	ligas=Espanha	0.2175	0.670	Continuar
3.º	odds	0.9971	0.670	<b>Parar</b>

Neste sentido o modelo final corresponde ao indicado na tabela 4.36.

Tabela 4.36: Modelo de Cox para 2.º Golo do Jogo de 1-0 para 1-1 com  $t_0 = t_{golo1}$ : modelo final

Covariável	$\beta_i$	valor-p
ligas=Alemanha	0.2288	0.0009
ligas=Inglaterra	0.1588	0.0157
ligas=Itália	0.2136	0.0008
odds	-0.0122	0.3247

#### 4.7.4.2 Funções Taxa de Falha e Taxa de Falha Acumulada

As imagens referentes a estas funções encontram-se na figura 4.37 e também na figura 4.38.

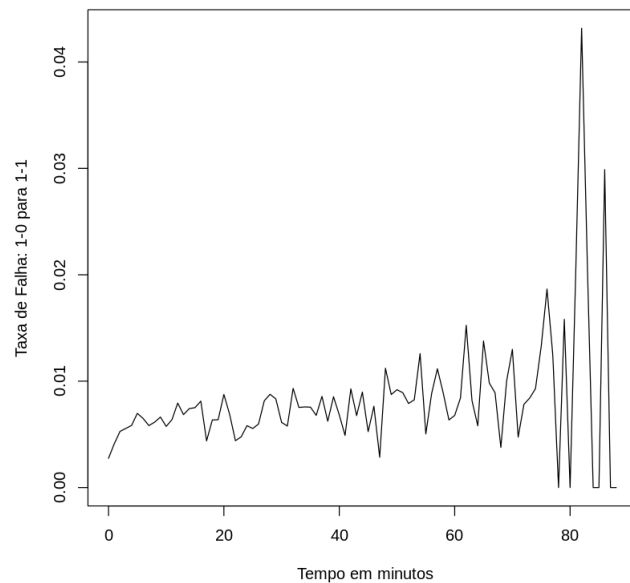


Figura 4.37: Função Taxa de Falha de 2.º Golo: de 1-0 para 1-1 com  $t_0 = t_{golo2}$

#### 4.7.5 Equipa de Fora: de 0-1 para 0-2

Nesta secção vamos apresentar os resultados referentes ao evento 2.º golo do jogo marcado por equipa de casa com  $t_0 = t_{golo1}$ , neste caso do 1-0 para o 2-0.

Salientamos que foram censurados todos os eventos de 2.º golo de jogo correspondentes aos casos: de 1-0 para 1-1, de 0-1 para 1-1 e de 1-0 para 2-0, também foram censurados os jogos em que ocorreu apenas 1 golo.

Iniciamos com o processo de seleção de variáveis e terminamos com a apresentação da taxa de falha e taxa de falha acumulada.



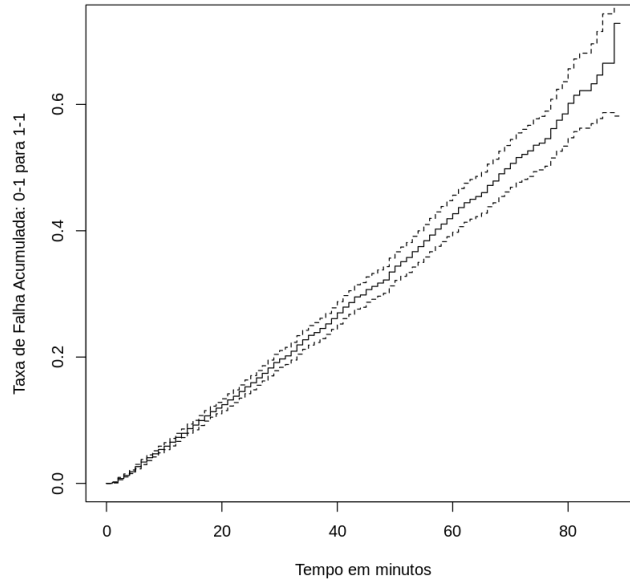


Figura 4.38: Função Taxa de Falha de 2.º Golo: de 1-0 para 1-1 com  $t_0 = t_{golo2}$

#### 4.7.5.1 Seleção de Variáveis

Para efeitos da construção deste modelo iniciamos com a inclusão das covariáveis indicadas na tabela 4.37.

Tabela 4.37: Modelo de Cox para 2.º Golo do Jogo de 0-1 para 0-2 com  $t_0 = t_{golo1}$ : modelo inicial

Covariável	$\beta_i$	valor-p
ligas=Espanha	0.2390	0.0145
ligas=Alemanha	0.3686	0.0003
ligas=Inglaterra	0.3995	$5 \times 10^{-5}$
ligas=França	0.3206	0.0015
ligas=Itália	0.3207	0.0014
odds	0.1747	$< 2 \times 10^{-16}$

Tal como realizado para outros eventos já analisados anteriormente executámos os passos indicados na tabela 4.38.

Tabela 4.38: Modelo de Cox para 2.º Golo do Jogo de 0-1 para 0-2 com  $t_0 = t_{golo1}$ : Passos de Seleção de Covariáveis

Passo	Covariável retirada	TRV	$\chi_p^2$	Decisão
1.º	ligas=Espanha	6.0547	0.670	<b>Parar</b>

Desta forma o modelo final corresponde ao modelo inicial.

#### 4.7.5.2 2.º Golo por Equipa de Casa: de 0-1 para 0-2

A figura 4.39 apresenta o gráfico da Função Taxa de Falha basal deste evento enquanto a figura 4.40 representa o gráfico da Função Taxa de Falha Acumulada.

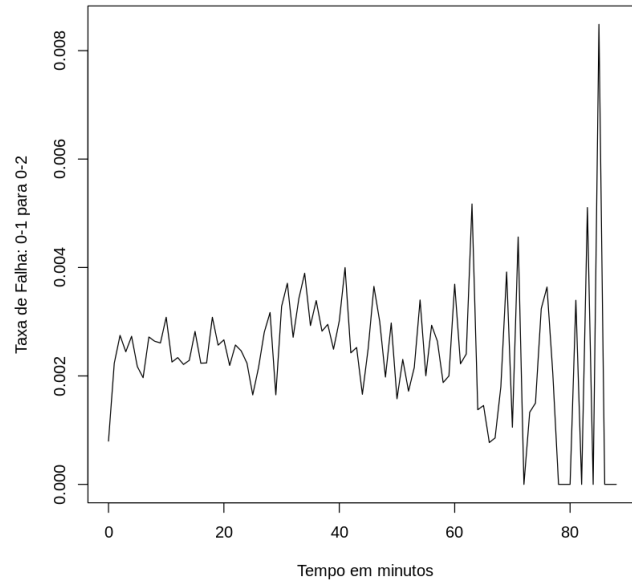


Figura 4.39: Função Taxa de Falha de 2.º Golo: de 0-1 para 0-2 com  $t_0 = t_{golo2}$

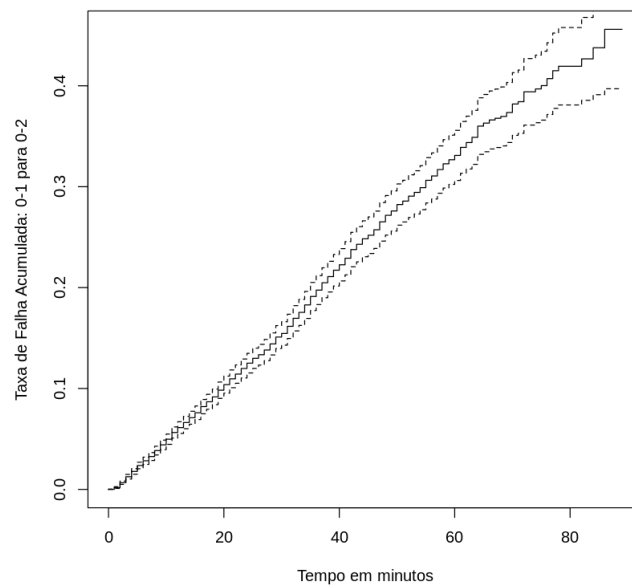


Figura 4.40: Função Taxa de Falha de 2.º Golo: de 0-1 para 0-2 com  $t_0 = t_{golo2}$

#### 4.7.6 Comparação de Funções Taxa de Falha: 1-0 para 2-0 vs. 1-0 para 1-1

A comparação das funções taxa de falha 1-0 para 2-0 vs. 1-0 para 1-1 encontra-se na figura 4.41.

Pode-se verificar que a taxa de falha do evento correspondente ao empate mantém-se constante ao longo do tempo, enquanto a taxa de falha referente ao 2.º golo da equipa de casa tem um período com valores muito semelhantes até aos 45-50 minutos e depois apresenta um crescimento relativamente acentuado. O que significa que a probabilidade instantânea de ocorrência do 2.º golo pela equipa de casa, tendo esta equipa marcado o 1.º golo, aumenta de forma notória a partir dos 45-50 minutos.

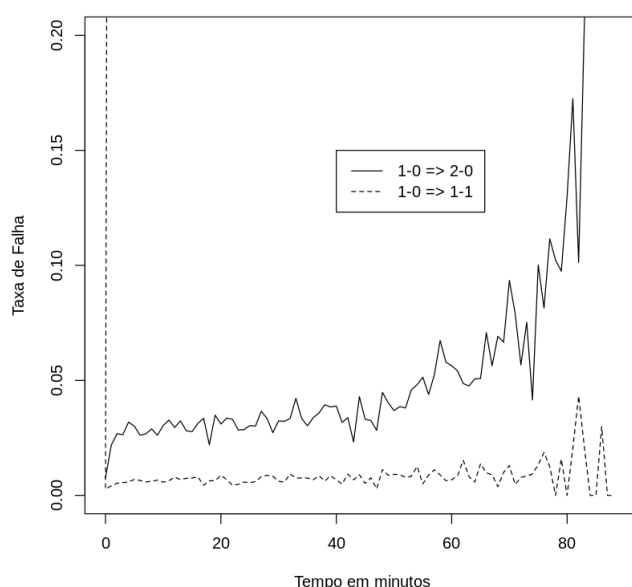


Figura 4.41: Comparação de Funções Taxa de Falha de 2.º Golo: de 1-0 para 2-0 vs. 1-0 para 1-1 com  $t_0 = t_{golo1}$

#### 4.7.7 Comparação de Funções Taxa de Falha: 0-1 para 0-2 vs. 0-1 para 1-1

A comparação de funções taxa de falha 0-1 para 0-2 vs. 0-1 para 1-1 encontra-se na figura 4.42.

É também, neste caso, evidente a preponderância do fator casa embora com muito maior variação da taxa de falha.

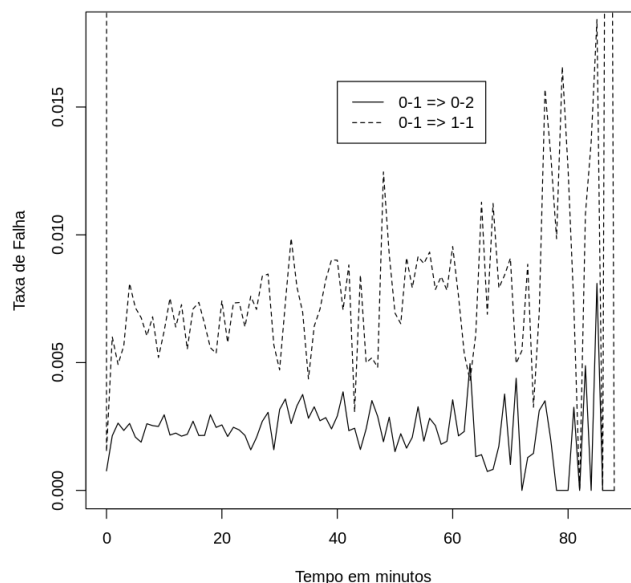


Figura 4.42: Comparação de Funções Taxa de Falha de 2.º Golo: de 0-1 para 0-2 vs. 0-1 para 1-1 com  $t_0 = t_{golo2}$

## 4.8 Estimativa de Kaplan-Meier para o 3.º Golo do Jogo

Nesta secção apresentamos os resultados referentes ao 3.º golo do jogo.

Iniciamos com a estimativa de Kaplan-Meier da função de sobrevivência do evento 3.º golo do jogo, prosseguimos com a estimativa da função de sobrevivência para dois tipos de eventos: 3.º golo marcado pela equipa de casa e 3.º golo marcado pela equipa de fora.

### 4.8.1 3.º Golo do Jogo

Para efeitos desta análise não foram incluídos os jogos em que só ocorreu um golo e foram censurados à direita os jogos em que ocorreram apenas dois golos.

Esta estimativa foi realizada considerando a variável ligas para permitir avaliar relativamente à igualdade das funções de sobrevivência de jogos provenientes de diversas ligas.

#### 4.8.1.1 Quartis de Tempo de Sobrevivência

Os quartis do tempo de sobrevivência para o evento 3.º golo do jogo foram os indicados na tabela 4.39.

Tabela 4.39: Quartis de Tempo de Sobrevivência para 3.º Golo do Jogo

Liga	1.º Quartil	Tempo Mediano	3.º Quartil
Portugal	61	82	ND
Espanha	55	78	ND
Alemanha	54	76	ND
Inglaterra	54.5	74	ND
França	60	82	ND
Itália	57	80	ND

ND - não disponível

#### 4.8.1.2 Função de Sobrevivência

Na figura 4.43 apresentamos os resultados do cálculo da estimativa de Kaplan-Meier para o referido evento.

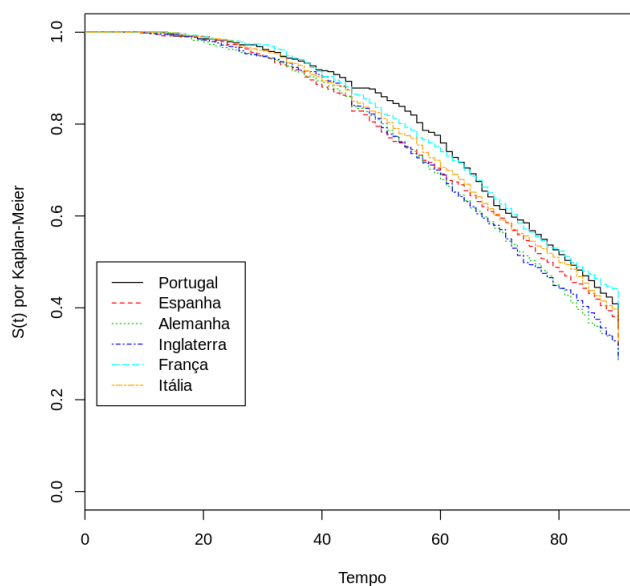


Figura 4.43: Estimativa de Kaplan-Meier da Função de Sobrevivência de 3.º Golo do Jogo

#### 4.8.2 3.º Golo Marcado pela Equipa de Casa

Para efeitos desta análise foram censurados os jogos em que ocorreu apenas um golo e os jogos em que a equipa de fora marcou o 2.º golo.

### 4.8.2.1 Quartis de Tempo de Sobrevivência

Os quartis obtidos para o tempo de sobrevivência deste evento foram os indicados na tabela 4.40.

Tabela 4.40: Quartis de Tempo de Sobrevivência para 3.º Golo do Jogo Marcado por Equipa de Casa

Liga	1.º Quartil	Tempo Mediano	3.º Quartil
Portugal	75	ND	ND
Espanha	68	ND	ND
Alemanha	66	90	ND
Inglaterra	65	ND	ND
França	70	ND	ND
Itália	69	ND	ND

ND - não disponível

### 4.8.2.2 Função de Sobrevivência

Na figura 4.44 apresentamos a estimativa de Kaplan-Meier para o 3.º golo do jogo marcado pela equipa de casa.

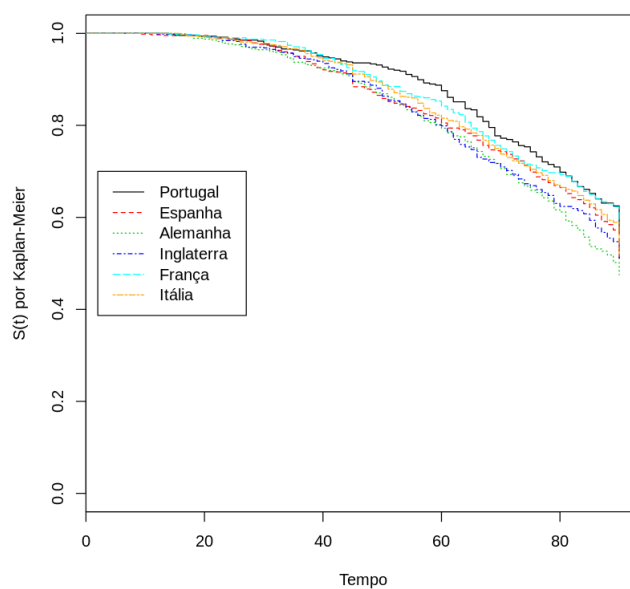


Figura 4.44: Estimativa de Kaplan-Meier da Função de Sobrevivência de 3.º Golo do Jogo Marcado por Equipa de Casa

### 4.8.3 3.º Golo do Jogo Marcado por Equipa de Fora

Para efeitos desta análise não foram incluídos os jogos em que ocorreu apenas um golo e foram censurados à direita os jogos em que se observaram apenas 2 golos e os jogos em que a equipa de fora marcou o 3.º golo.

#### 4.8.3.1 Quartis de Tempo de Sobrevivência

Os quartis obtidos para o tempo de sobrevivência deste evento foram os indicados na tabela 4.41.

Tabela 4.41: Quartis de Tempo de Sobrevivência para 3.º Golo do Jogo Marcado por Equipa de Fora

Liga	1.º Quartil	Tempo Mediano	3.º Quartil
Portugal	79	ND	ND
Espanha	77	ND	ND
Alemanha	78	ND	ND
Inglaterra	74	ND	ND
França	81	ND	ND
Itália	81	ND	ND

ND - não disponível

#### 4.8.3.2 Função de Sobrevivência

Na figura 4.45 apresentamos a estimativa de Kaplan-Meier para o 3.º golo do jogo marcado pela equipa de fora.

## 4.9 Modelo de Cox para o 3.º Golo do Jogo

Nesta secção vamos apresentar os resultados obtidos com o ajustamento do modelo de Cox aos dados referentes ao 3.º golo do jogo. Iniciamos com a verificação da hipótese de riscos proporcionais, prosseguimos com o processo de seleção de variáveis e completamos com as funções taxa de falha e taxa de falha acumulada.

### 4.9.1 Hipótese de Riscos Proporcionais

Para análise dos resíduos de Schoenfeld foi produzida a figura 4.46.

Em face dos resultados constatamos que a variável casa não pode ser incluída no modelo de Cox o que indica que, para uma melhor análise dos resultados será mais adequado definir dois modelos diferentes em função da equipa que marcou o golo: casa ou fora.

Isto significa que uma melhor análise corresponderá a considerar os diversos eventos que estão associados ao 3.º golo.

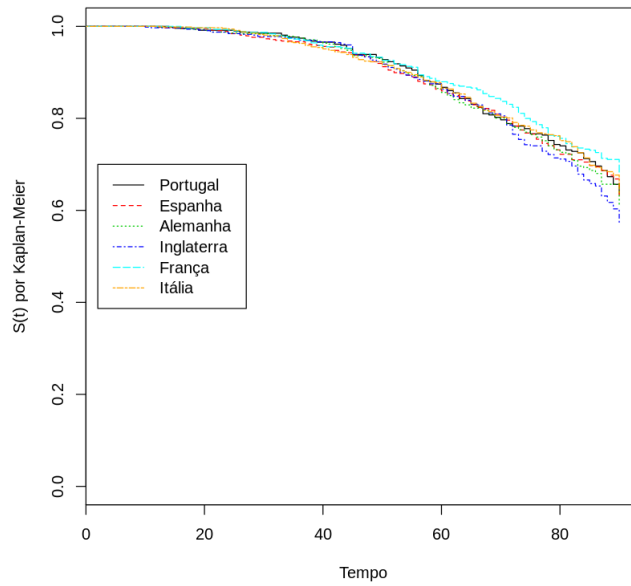


Figura 4.45: Estimativa de Kaplan-Meier da Função de Sobrevivência 3.º Golo Marcado por Equipa de Fora

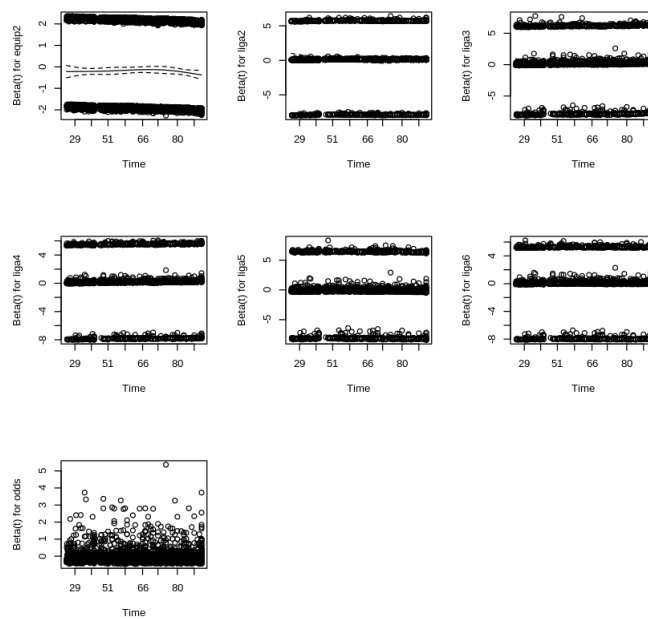


Figura 4.46: Resíduos de Schoenfeld de 3.º Golo do Jogo

Existindo 8 tipos de eventos alguns terão um significado mais relevante na marcha do marcador para o desenho do resultado final.

Neste sentido escolhemos dois eventos que correspondem ao desempate no marcador, sendo um alcançado pela equipa de casa: de 1-1 para 2-1, e o outro alcançado pela equipa de fora: de 1-1 para 1-2.



São estes os eventos que iremos analisar nas subsecções seguintes.

#### 4.9.2 Equipa de Casa: de 1-1 para 2-1

Nesta secção vamos apresentar os resultados referentes ao evento 3.º golo marcado por equipa de casa, neste caso do evento correspondente à passagem de 1-1 para 2-1.

Salientamos que foram censurados todos os eventos de 3.º golo do jogo correspondentes aos casos:

- de 2-0 para 3-0,
- de 2-0 para 2-1,
- de 0-2 para 0-3,
- de 0-2 para 1-2,

também foram censurados os jogos em que ocorreram apenas 2 golos.

Iniciamos com o processo de seleção de variáveis e completamos com as funções taxa de falha e taxa de falha acumulada.

##### 4.9.2.1 Seleção de Variáveis

Para efeitos desta análise iniciamos com a inclusão das covariáveis indicadas na tabela 4.42.

Tabela 4.42: Modelo de Cox para 3.º Golo do Jogo: de 1-1 para 2-1 - modelo inicial

Covariável	$\beta_i$	valor-p
ligas=Espanha	0.2056	0.0387
ligas=Alemanha	0.3159	0.0017
ligas=Inglaterra	0.2369	0.0173
ligas=França	-0.0104	0.9198
ligas=Itália	0.1635	0.0961
odds	-0.0032	0.7975

Tal como realizado para outros eventos já analisados executámos os passos indicados na tabela 4.43.

Tabela 4.43: Modelo de Cox para 3.º Golo do Jogo: de 1-1 para 2-1 - Passos de Seleção de Covariáveis

Passo	Covariável retirada	TRV	$\chi_p^2$	Decisão
1.º	ligas=França	0.0101	0.670	Continuar
2.º	odds	0.0631	0.670	Continuar
3.º	ligas=Itália	4.5678	0.670	<b>Parar</b>

Neste sentido o modelo final corresponde ao indicado na tabela 4.44.

Tabela 4.44: Modelo de Cox para 3.º Golo do Jogo: de 1-1 para 2-1 - modelo final

Covariável	$\beta_i$	valor-p
ligas=Espanha	0.2114	0.0087
ligas=Alemanha	0.3218	$8 \times 10^{-5}$
ligas=Inglaterra	0.2428	0.0026
ligas=Itália	0.1694	0.0320

#### 4.9.2.2 Funções Taxa de Falha e Taxa de Falha Acumulada

A função taxa de falha deste evento encontra-se na figura 4.47.

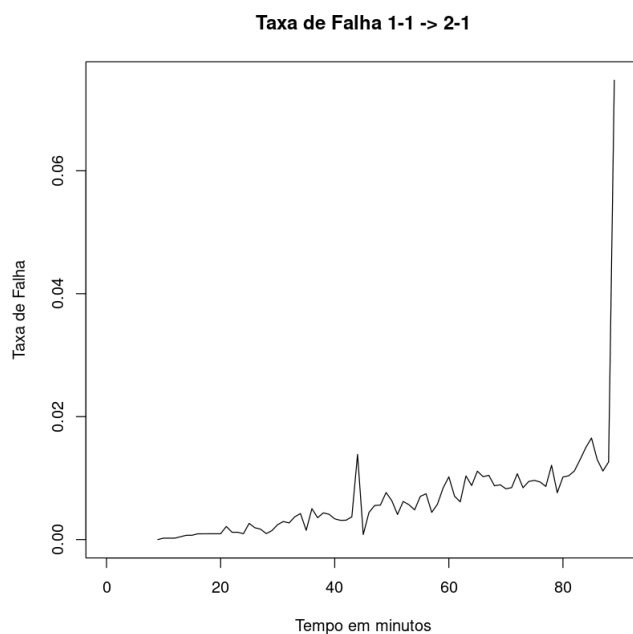


Figura 4.47: Função Taxa de Falha de 3.º Golo: de 1-1 para 2-1

A função taxa de falha acumulada deste evento encontra-se na figura 4.48.

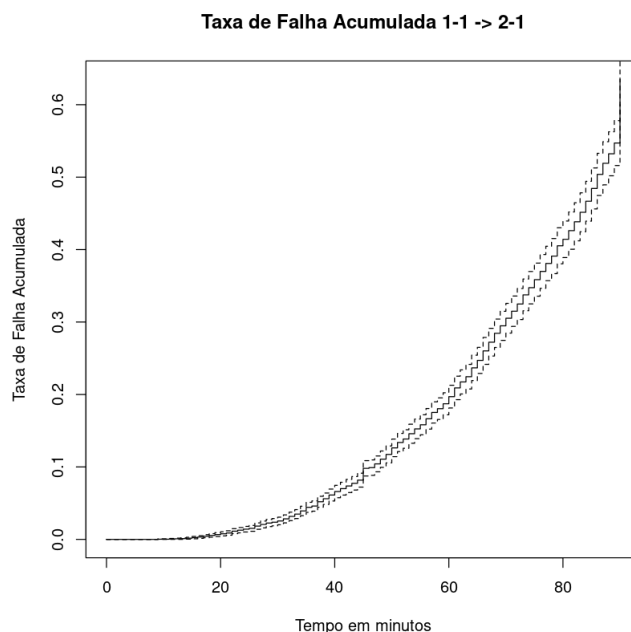


Figura 4.48: Função Taxa de Falha Acumulada de 3.º Golo: de 1-1 para 2-1

### 4.9.3 Equipa de Fora: de 1-1 para 1-2

Nesta secção vamos apresentar os resultados referentes ao evento 3.º golo marcado por equipa de fora, neste caso de 1-1 para 1-2.

Salientamos que foram censurados todos os eventos de 3.º golo do jogo correspondentes aos casos:

- de 2-0 para 3-0,
- de 2-0 para 2-1,
- de 0-2 para 0-3,
- de 0-2 para 1-2,

também foram censurados os jogos em que ocorreram apenas 2 golos.

Iniciamos pelo processo de seleção de variáveis e finalizamos com a apresentação da taxa de falha e taxa de falha acumulada.

#### 4.9.3.1 Seleção de Variáveis

Para efeitos desta análise iniciamos com a inclusão das covariáveis indicadas na tabela 4.45.

Tal como realizado para outros eventos já analisados executámos os passos indicados na tabela 4.46.

Neste sentido o modelo final corresponde ao indicado na tabela 4.47.

Tabela 4.45: Modelo de Cox para 3.º Golo do Jogo: de 1-1 para 1-2 - modelo inicial

Covariável	$\beta_i$	valor-p
ligas=Espanha	0.0262	0.816
ligas=Alemanha	0.0621	0.593
ligas=Inglaterra	0.1617	0.144
ligas=França	-0.1568	0.179
ligas=Itália	-0.0119	0.915
odds	-0.0171	0.272

Tabela 4.46: Modelo de Cox para 3.º Golo do Jogo: de 1-1 para 1-2 - Passos de Seleção de Covariáveis

Passo	Covariável retirada	TRV	$\chi_p^2$	Decisão
1.º	ligas=Itália	0.0114	0.670	Continuar
2.º	ligas=Espanha	0.1277	0.670	Continuar
3.º	ligas=Alemanha	0.3970	0.670	Continuar
4.º	odds	1.2966	0.670	<b>Parar</b>

Tabela 4.47: Modelo de Cox para 3.º Golo do Jogo: de 1-1 para 1-2 - modelo final

Covariável	$\beta_i$	valor-p
ligas=Inglaterra	0.1438	0.0749
ligas=França	0.1747	0.049
odds	0.0173	0.2658

#### 4.9.3.2 Funções Taxa de Falha e Taxa de Falha Acumulada

A função taxa de falha deste evento encontra-se na figura 4.49.

A função taxa de falha acumulada deste evento encontra-se na figura 4.50.

#### 4.9.4 Comparação das Funções Taxa de Falha de 1-1 para 2-1 vs. de 1-1 para 1-2

Terminamos esta secção referente ao 3.º golo do jogo com a apresentação das funções taxa de falha que se pode ver na figura 4.51.

Como se pode verificar não se podem indicar grandes diferenças entre estes dois tipos de eventos.

### 4.10 Estimativa de Kaplan-Meier do 3.º Golo do Jogo com Tempo Inicial Igual ao Tempo do 2.º Golo

Nesta secção apresentamos os resultados referentes ao 3.º golo do jogo. Iniciamos com a estimativa de Kaplan-Meier da função de sobrevivência de três tipos de eventos:

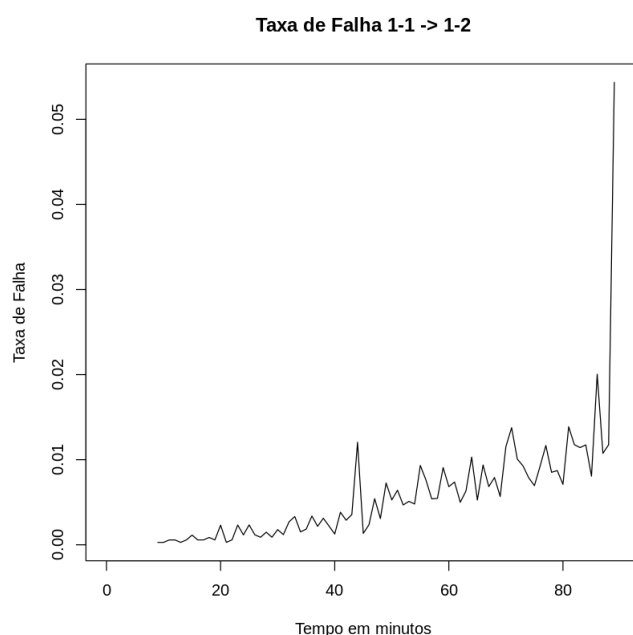


Figura 4.49: Função Taxa de Falha de 3.º Golo: de 1-1 para 1-2

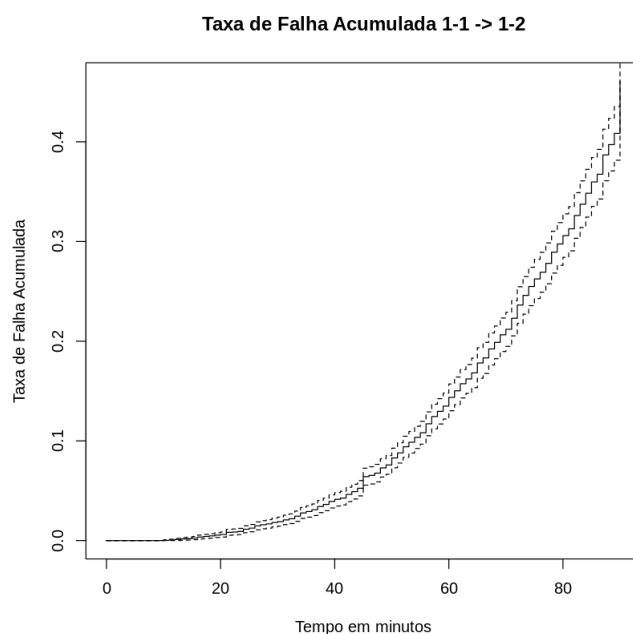


Figura 4.50: Função Taxa de Falha Acumulada de 3.º Golo: de 1-1 para 1-2

3.º golo do jogo, 3.º golo do jogo marcado pela equipa de casa e 3.º golo do jogo marcado pela equipa de fora. Saliente-se que para efeitos das análises referentes ao 3.º golo são considerados os jogos em que ocorreram pelo menos dois golos, os únicos jogos que se pode dizer terem atingido um estado que reunia condições para a ocorrência deste evento (3.º golo).

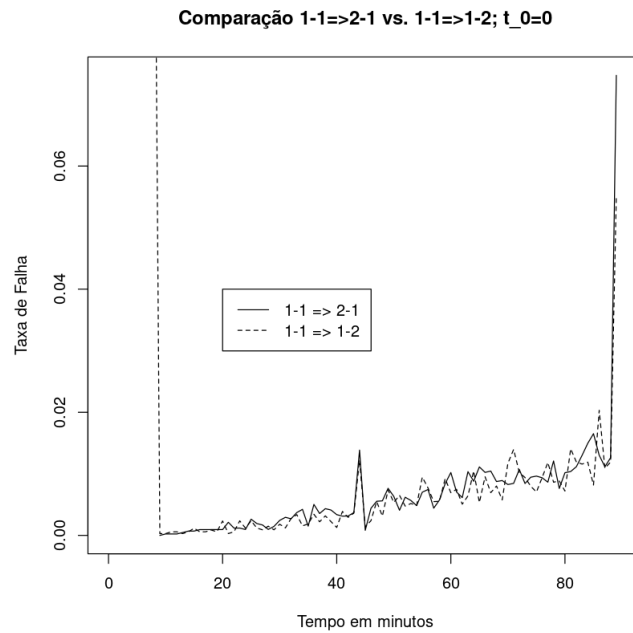


Figura 4.51: Comparação das Funções Taxa de Falha de 1-1 para 2-1 vs. de 1-1 para 1-2

### 4.10.1 3.º Golo do Jogo

Para efeitos desta análise foram censurados os jogos em que ocorreu apenas um golo e os jogos em que a equipa de fora marcou o 2.º golo, o tempo do evento foi obtido subtraindo ao tempo efetivo o tempo registado para o 2.º golo do jogo.

Os quartis do tempo de sobrevivência obtidos para este evento foram os indicados na tabela 4.48.

Tabela 4.48: Quartis de Tempo de Sobrevivência para 3.º Golo do Jogo com  $t_0 = t_{golo2}$

Liga	1.º Quartil	Tempo Mediano	3.º Quartil
Portugal	12	25	49
Espanha	9	22	42
Alemanha	10	24	44
Inglaterra	9	20	41
França	11	23	46
Itália	9	21	42

#### 4.10.1.1 Função de Sobrevivência

Na figura 4.52 apresentam-se os resultados do cálculo da estimativa de Kaplan-Meier para as diversas ligas relativo ao 3.º golo com  $t_0 = t_{golo2}$ .

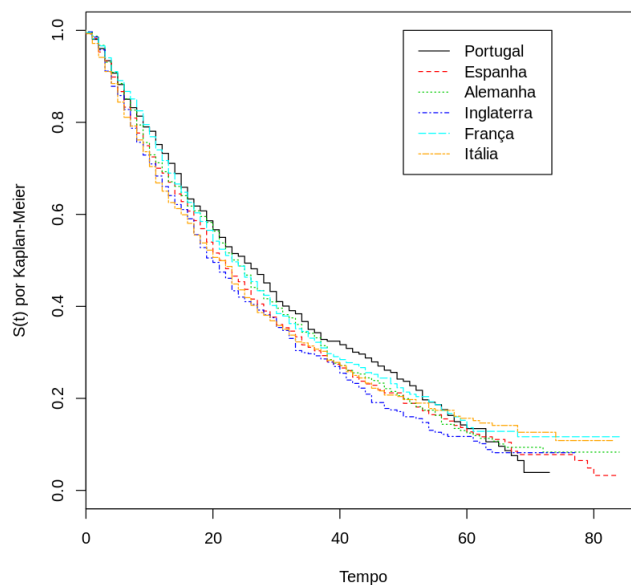


Figura 4.52: Estimativa de Kaplan-Meier da Função de Sobrevivência 3.º Golo do Jogo com  $t_0 = t_{golo2}$

### 4.10.2 3.º Golo Marcado por Equipa de Casa

De seguida apresentamos os resultados da estimativa de Kaplan-Meier para o referido evento. Iniciamos com a indicação dos quartis do tempo de sobrevivência e posteriormente passamos à função de sobrevivência.

#### 4.10.2.1 Quartis de Tempo de Sobrevivência

Os quartis do tempo de sobrevivência obtidos para este evento foram os indicados na tabela 4.49.

Tabela 4.49: Quartis de Tempo de Sobrevivência para 3.º Golo do Jogo Marcado por Equipa de Casa com  $t_0 = t_{golo2}$

Liga	1.º Quartil	Tempo Mediano	3.º Quartil
Portugal	19	48	68
Espanha	16	36	66
Alemanha	16	38	65
Inglaterra	15	37	ND
França	17	38	ND
Itália	15	36	ND

ND - não disponível

### 4.10.2 Função de Sobrevivência

Na figura 4.53 apresentam-se os resultados do cálculo da estimativa de Kaplan-Meier para as diversas ligas relativo ao 3.º golo com  $t_0 = t_{golo2}$ .

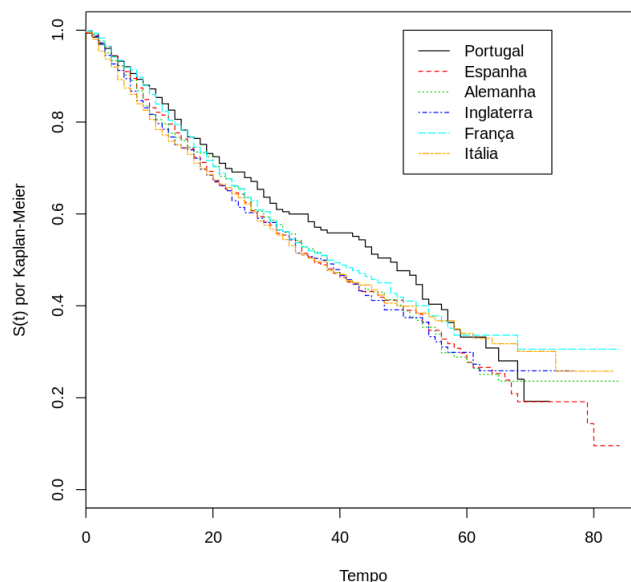


Figura 4.53: Estimativa de Kaplan-Meier da Função de Sobrevivência 3.º Golo do Jogo Marcado por Equipa de Casa com  $t_0 = t_{golo2}$

### 4.10.3 3.º Golo Marcado por Equipa de Fora

Em seguida apresentamos os resultados da estimativa de Kaplan-Meier para o referido evento. Iniciamos com a indicação dos quartis do tempo de sobrevivência e posteriormente passamos à função de sobrevivência.

Os quartis do tempo de sobrevivência obtidos para este evento foram os indicados na tabela 4.50.

Tabela 4.50: Quartis de Tempo de Sobrevivência para 3.º Golo do Jogo Marcado por Equipa de Casa com  $t_0 = t_{golo2}$

Liga	1.º Quartil	Tempo Mediano	3.º Quartil
Portugal	23	51	69
Espanha	22	50	ND
Alemanha	24	55	ND
Inglaterra	19	45	ND
França	29	55	ND
Itália	22	54	ND

ND - não disponível



### 4.10.3.1 Função de Sobrevivência

Na figura 4.54 apresentam-se os resultados do cálculo da estimativa de Kaplan-Meier para as diversas ligas relativo ao 3.º golo com  $t_0 = t_{golo2}$ .

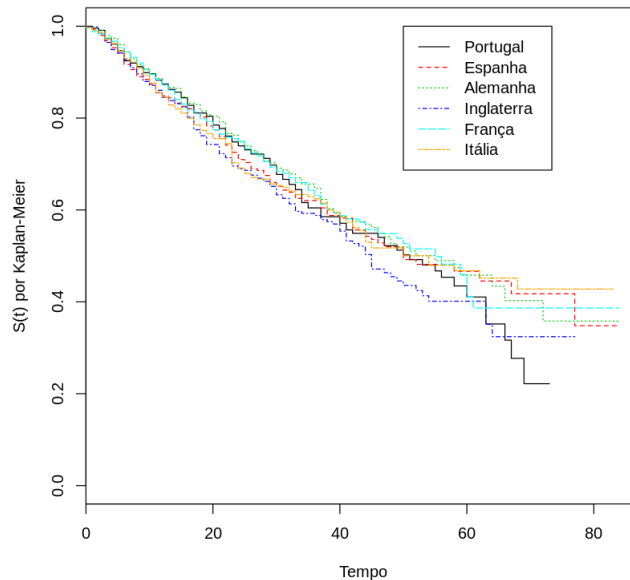


Figura 4.54: Estimativa de Kaplan-Meier da Função de Sobrevivência 3.º Golo do Jogo Marcado por Equipa de Fora com  $t_0 = t_{golo2}$

## 4.11 Modelo de Cox para o 3.º Golo do Jogo com Tempo Inicial Igual a Tempo do 2.º Golo

Nesta secção vamos apresentar os resultados obtidos a partir do ajustamento do modelo de Cox aos dados referentes ao 3.º golo do jogo, desta feita com com momento inicial sendo  $t_0 = t_{golo2}$ .

### 4.11.1 Hipótese de Riscos Proporcionais

Para avaliar da hipótese de riscos proporcionais foi produzido o gráfico da figura 4.55. Este gráfico indica que, à semelhança dos resultados obtidos para o mesmo evento considerado em função do início do jogo, a maioria das variáveis respeita a hipótese de riscos proporcionais com exceção da variável equipa (casa/fora).

Considerando esta conclusão, e também para permitir a comparação com os resultados obtidos na secção anterior, decidimos estudar os mesmos dois tipos de eventos: do 1-1 para 2-1 e do 1-1 para 1-2.

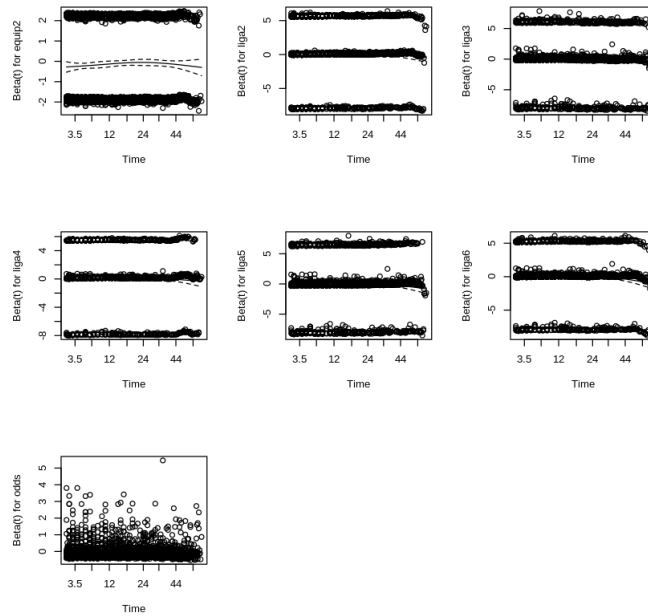


Figura 4.55: Resíduos de Schoenfeld para 3.º Golo do Jogo com  $t_0 = t_{golo2}$

## 4.11.2 Equipa de Casa: de 1-1 para 2-1

Nesta subsecção vamos proceder à análise de sobrevivência do evento 3.º golo do jogo marcado por equipa de casa, considerando nomeadamente o evento que corresponde ao desempate: de 1-1 para o 2-1.

### 4.11.2.1 Seleção de Variáveis

Para efeitos desta análise iniciamos com a inclusão das covariáveis indicadas na tabela 4.51.

Tabela 4.51: Modelo de Cox para 3.º Golo do Jogo: de 1-1 para 2-1 com  $t_0 = t_{golo2}$  - modelo inicial

Covariável	$\beta_i$	valor-p
ligas=Espanha	0.1823	0.067
ligas=Alemanha	0.1624	0.106
ligas=Inglaterra	0.2001	0.0442
ligas=França	0.0720	0.485
ligas=Itália	0.1736	0.077
odds	$1.99 \times 10^{-6}$	0.999

Tal como realizado para outros eventos já analisados executámos os passos indicados na tabela 4.52.

Tabela 4.52: Modelo de Cox para 3.º Golo do Jogo: de 1-1 para 2-1 com  $t_0 = t_{golo2}$  - Passos de Seleção de Covariáveis

Passo	Covariável retirada	TRV	$\chi_p^2$	Decisão
1.º	odds	2.49	0.670	<b>Parar</b>

#### 4.11.2.2 Funções Taxa de Falha e Taxa de Falha Acumulada

A função taxa de falha deste evento corresponde ao gráfico indicado na figura 4.56.

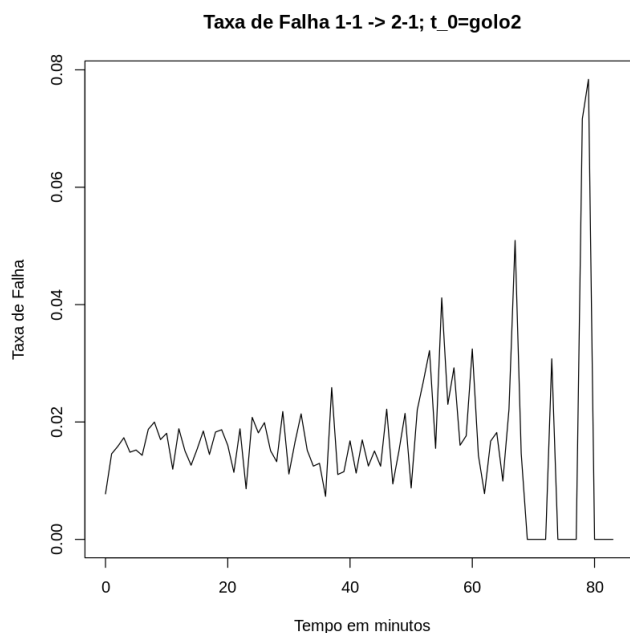


Figura 4.56: Função Taxa de Falha de 3.º Golo: de 1-1 para 2-1 com  $t_0 = t_{golo2}$

A função taxa de falha acumulada deste evento corresponde ao gráfico indicado na figura 4.57.

#### 4.11.3 Equipa de Fora: de 1-1 para 1-2 com $t_0 = t_{golo2}$

Nesta subsecção vamos proceder à análise de sobrevivência do evento 3.º golo do jogo marcado por equipa de casa, considerando nomeadamente o evento que corresponde ao desempate: de 1-1 para o 1-2.

##### 4.11.3.1 Seleção de Variáveis

Para efeitos desta análise iniciamos com a inclusão das covariáveis indicadas na tabela 4.53.

Tal como realizado para outros eventos já analisados executámos os passos indicados na tabela 4.54.

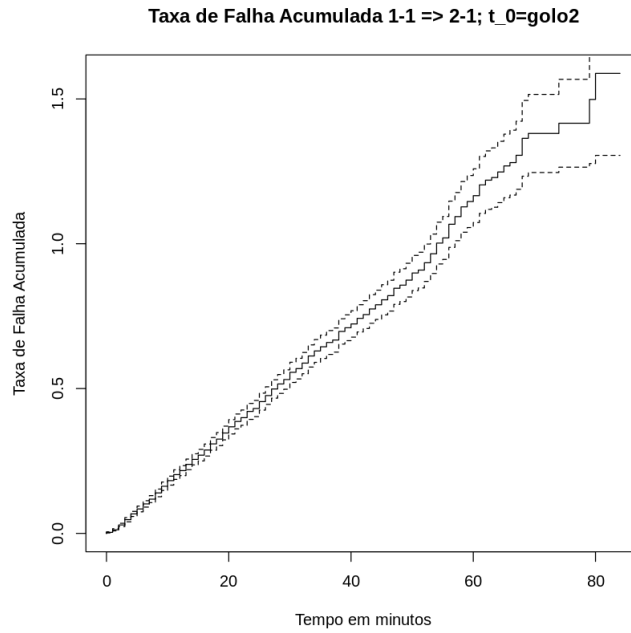


Figura 4.57: Função Taxa de Falha de 3.º Golo: de 1-1 para 2-1 com  $t_0 = t_{golo2}$

Tabela 4.53: Modelo de Cox para 3.º Golo do Jogo: de 1-1 para 1-2 com  $t_0 = t_{golo2}$  - modelo inicial

Covariável	$\beta_i$	valor-p
ligas=Espanha	0.0015	0.990
ligas=Alemanha	-0.0978	0.401
ligas=Inglaterra	0.1192	0.281
ligas=França	-0.7409	0.525
ligas=Itália	-0.0021	0.985
odds	-0.0138	0.376

Tabela 4.54: Modelo de Cox para 3.º Golo do Jogo: de 1-1 para 1-2 com  $t_0 = t_{golo2}$  - Passos de Seleção de Covariáveis

Passo	Covariável retirada	TRV	$\chi_p^2$	Decisão
1.º	ligas=Espanha	7.8747	0.378	<b>Parar</b>

Neste sentido o modelo final corresponde ao indicado na tabela 4.53 que representa os coeficientes do modelo inicial.

#### 4.11.3.2 Funções Taxa de Falha e Taxa de Falha Acumulada

A função taxa de falha deste evento corresponde ao indicado na figura 4.58.

A função taxa de falha acumulada para este evento corresponde à indicada na figura 4.59.

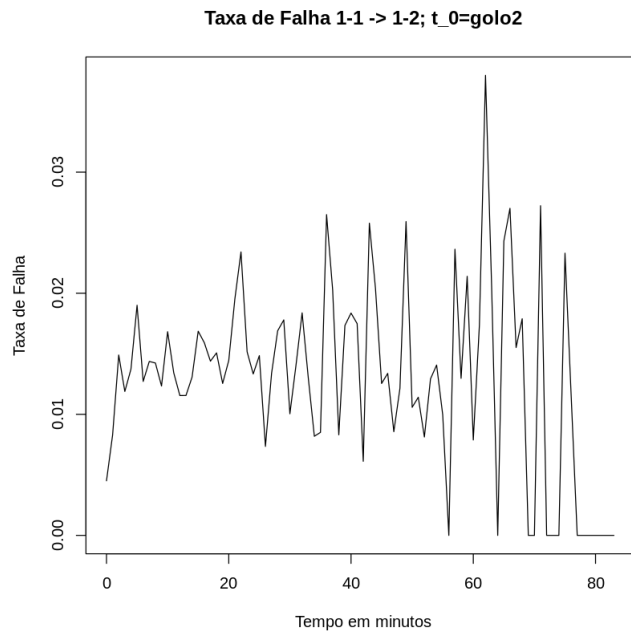


Figura 4.58: Função Taxa de Falha de 3.º Golo: de 1-1 para 1-2 com  $t_0 = t_{golo2}$

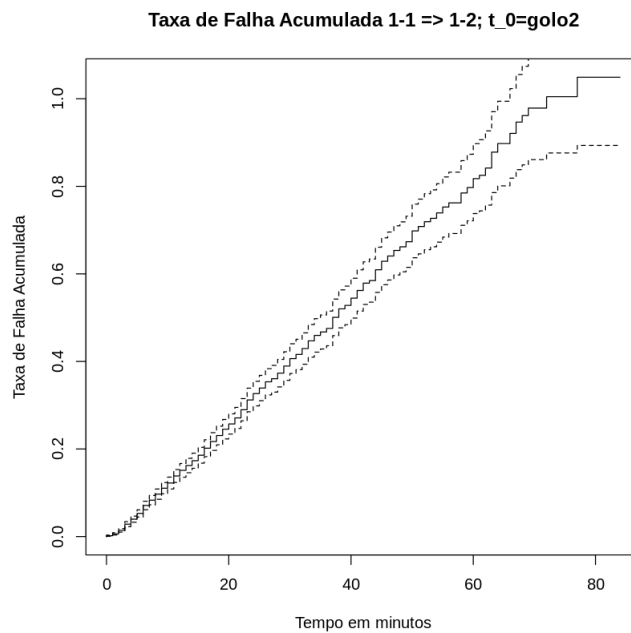


Figura 4.59: Função Taxa de Falha de 3.º Golo: de 1-1 para 1-2 com  $t_0 = t_{golo2}$

#### 4.11.4 Comparação de Função Taxa de Falha de 1-2 para 2-1 vs. 1-1 para 1-2 com $t_0 = t_{golo2}$

Esta última parte destina-se a apresentar a comparação das funções taxa de falha do evento de 1-2 para 2-1 com o evento de 1-1 para 1-2 com  $t_0 = t_{golo2}$  e que se encontra na figura 4.60.

Como se pode verificar não existe uma grande diferença na taxa de falha dos dois tipos de evento, indicando neste caso que o fator casa não tem a mesma preponderância que foi revelada noutros eventos de golo de ordem inferior.

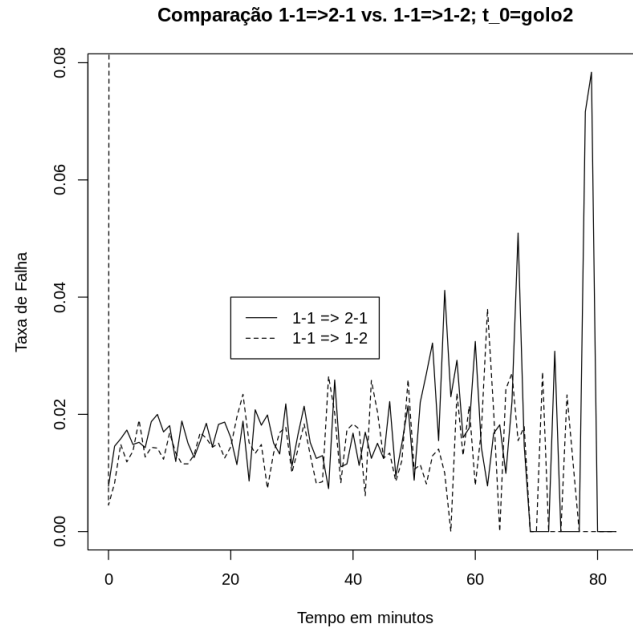


Figura 4.60: Comparação de Funções Taxa de Falha de 3.º Golo: 1-1 para 1-2 vs. 1-2 para 2-1 com  $t_0 = t_{golo2}$

# Capítulo 5

## Conclusões

O principal objetivo deste estudo foi caracterizar e comparar a sobrevivência de eventos de golo em jogos de futebol. Pretendeu-se também identificar e caracterizar a forma como diversas variáveis afetavam o comportamento da função de sobrevivência destes eventos.

Para esse efeito foram aplicados diversos procedimentos de recolha, armazenamento e análise de dados, procedimentos estes que nos permitiram a concretização dos objetivos relativos a diversos eventos de golo que identificámos: 1.º golo, 2.º golo e 3.º golo do jogo.

A facilidade de acesso aos dados, por estarem disponíveis em documentos online, constituiu um elemento favorável à concretização do estudo. Porém, outros fatores condicionaram a aplicação da metodologia de análise de sobrevivência em particular por não haver forma de recolher informações sobre algumas variáveis que poderiam influenciar a definição da função de sobrevivência.

Assim, variáveis como a estabilidade da equipa, classificação no campeonato, substituições (covariável dependente do tempo), cartões (amarelos, vermelhos) e outras, não foram incluídas neste estudo quer por questões de tempo de análise quer por não estarem acessíveis de forma expedita.

Nas secções seguintes passamos a apresentar de forma resumida as principais conclusões referentes a cada um dos tipos de eventos.

### 5.1 1.º Golo

Relativamente ao 1.º golo tornou-se desde logo evidente, nos resultados obtidos, o que diversos autores concluíram em seus estudos: o facto do fator casa ser de importância relevante para determinar o resultado do jogo.

Também na análise de sobrevivência deste evento se concluiu que existe diferença entre equipa de casa marca o seu 1.º golo e equipa de fora marca o seu 1.º golo, com evidente maior probabilidade para o primeiro destes, resultando em funções de sobrevivência e tempos de sobrevivência médios com valores inferiores.

Assim, verificou-se que a probabilidade de sobrevivência no final dos 90 minutos, obtida por via do estimador de Kaplan-Meier, para o evento equipa de casa marca o seu 1.º golo foi de 0.25-0.30 e tempos medianos em torno dos 45-50 minutos, enquanto para o evento equipa de fora marca o seu 1.º golo foi de 0.35-0.40 e tempos medianos

em geral em torno dos 60 minutos.

Também o mesmo se verificou para o 1.º golo do jogo, embora desta feita com valores diferentes.

Na construção do modelo de Cox para este evento constatou-se que a variável não respeita a hipótese de riscos proporcionais o que nos levou a definir dois modelos diferentes para os dois tipos de equipas: casa e fora.

No primeiro caso (modelo casa) o modelo final após o processo de seleção de variáveis contém 3 covariáveis sendo estas: *odds*, liga=Alemanha e liga=França.

No segundo caso (modelo fora) o modelo final contém 4 covariáveis: *odds*, liga=Espanha, liga=Alemanha e liga=Itália.

Saliente-se que no primeiro caso o coeficiente da variável *odds* tem valor negativo e no segundo caso o coeficiente tem valor positivo, o que faz sentido pois, em princípio, quanto maior o valor de *odds* maior a probabilidade de sobrevivência ao evento 1.º golo marcado por equipa de fora, e inversamente para o primeiro caso.

Por fim, a função taxa de falha é sempre superior ao longo do tempo e também que a liga da Alemanha tem uma função de sobrevivência diferente das restantes, pois a covariável está presente nos dois modelos.

## 5.2 2.º Golo

Quanto ao 2.º golo do jogo também se verificou a preponderância do fator casa, com tempos medianos de sobrevivência em torno 80 minutos embora as probabilidade de sobrevivência aos 90 minutos serem muito semelhantes, entre 0.45-0.50.

Igualmente neste caso, no ajustamento dos dados ao modelo de Cox também a variável equipa (casa/fora) se verificou não respeitar a hipótese de riscos proporcionais, pelo que se adotou uma metodologia semelhante à que já tinha sido adotada para o 1.º golo do jogo, que consistiu em identificar eventos em função da equipa que marcou e construir modelos em função do tipo de evento de golo.

Desta forma foram identificados diversos eventos de golo em função da equipa que o marcou donde se constataram diversas diferenças no processo de seleção de variáveis.

Também procedemos à análise de sobrevivência deste evento considerando como tempo inicial o tempo referente ao 1.º golo do jogo.

Neste caso os tempos medianos de sobrevivência rondaram os 25 min. com 3.º quartil em torno dos 45-50 minutos.

Em suma, o que se constatou na comparação dos eventos correspondentes ao 2.º golo do jogo, usando como instante inicial o tempo do 1.º golo, foi que a taxa de falha do golo de 1-0 para 2-0 apresentava uma tendência de crescimento a partir do minuto 45/50, a contar do instante do 1.º golo, mantendo-se com um valor muito semelhante no período anterior a este.

Por outro lado, para a comparação dos eventos de golo de 0-1 para 0-2 vs. de 0-1 para 1-1, quando se definiu como instante inicial o começo do jogo, verificou-se que a taxa de falha era superior no 1.º destes e os dois apresentavam, para aquela função, uma bem definida tendência de crescimento ao longo do tempo de jogo. Porém, quando se considerou o instante inicial como o tempo de ocorrência do 1.º golo do jogo, a taxa de falha do mesmo evento não apresentava uma tendência de crescimento tão bem definida.



As diferenças indicadas nos dois parágrafos anteriores são então mais notórias quando se observa o que se afirma no primeiro destes parágrafos, o que nos leva a concluir, de forma diferente do que foi concluído pelos autores (*Nevo e Ritov, 2012*), a ocorrência do 1.º golo de facto induz a ocorrência do 2.º golo quando se trata do 1.º golo marcado pela equipa de casa (o 1-0) mas induz de forma diferida por um período 45-50 minutos a partir do qual a respetiva taxa de falha tem um crescimento mais acentuado.

Este facto deverá ocorrer talvez devido ao cansaço provocado na equipa visitante pelo esforço de modificar o rumo dos acontecimentos ou de evitar a repetição de um evento de golo pela equipa de casa.

Também neste caso se verificou a importância do fator casa que se revelou no facto das taxas de falha basal serem sempre maiores no caso do evento de 1-0 para 2-0 e 0-1 para 1-1, do que respetivamente de 1-0 para 1-1 e 0-1 para 0-2.

### 5.3 3.º Golo

Quanto a este evento constatou-se desde logo que os quartis de tempo de sobrevivência são naturalmente mais elevados (tempo mediano em torno dos 75-80 min.) o que levou também a obter taxas de falha muito mais reduzidas (no caso do evento de golo de 1-1 para 2-1 foi em geral inferior a 0.02).

A nossa análise concentrou-se nos eventos de golo correspondentes ao desempate e o que se verificou foi que na comparação das taxas de falha dos dois tipos de evento (de 1-1 para 2-1 vs. de 1-2 para 1-2) se observou uma tendência de crescimento muito semelhante e sem qualquer preponderância do fator casa quando se considerou o instante inicial correspondente ao começo do jogo.

Por outro lado, quando se considerou o instante inicial correspondente ao tempo do 2.º golo do jogo ( $t_0 = t_{golo2}$ ) a função taxa de falha não revelou qualquer tendência de crescimento em qualquer dos dois tipos de evento.

Em qualquer caso, o fato de aqui não se verificar preponderância do fator casa parece indicar que nestes casos os jogos são mais equilibrados e nenhuma equipa parece ter preponderância na definição do rumo dos eventos.

### 5.4 Perspetivas de Trabalho de Futuro

Nesta secção vamos indicar algumas orientações de desenvolvimento de trabalho futuro considerando quer os resultados obtidos quer as limitações que estiveram na elaboração deste estudo. Uma primeira observação vai para o conjunto de covariáveis que decorreram da recolha de dados. Em abordagens futuras este estudo poderá ser enriquecido pela inclusão de mais variáveis de que podemos apresentar como exemplos: a classificação no campeonato à data do jogo, estabilidade da equipa/plantel, cartões, substituições (covariável dependente do tempo), o que implica o uso de modelos com variáveis dependentes do tempo.

Por outro lado, tal como para a maioria dos estudos estatísticos as abordagens futuras poderão incluir um maior número de épocas desportivas e recorrer a mais ligas desportivas (europeias e não europeias).

Na vertente de técnicas estatísticas, abordagens futuras poderão recorrer à extensão do modelo de Cox com inclusão de uma variável fragilidade que seguirá uma determinada distribuição e se destinará a modelar o efeito de uma covariável não observável (*Rocha e Papoila, 2008*).

# Bibliografia

- P. Serranho. *Notas de Análise de Sobrevivência - Uma Introdução com R*. Universidade Aberta, Lisboa, 2015
- Project Jupyter. Project jupyter - home. URL <https://jupyter.org/>. [acedido em maio-2018]
- G. L. Silva e M. A. A. Turkman. *Modelos Lineares Generalizados - da teoria à prática*. Sociedade Portuguesa de Estatística, Lisboa, 2000.
- C. Rocha e A. L. Papoila. *Análise de Sobrevivência*. Sociedade Portuguesa de Estatística, Lisboa, 2009.
- J. P. Klein, H. C. van Houwelingen, J. G. Ibrahim e T. H. Scheike. *Handbook of Survival Analysis*. CRC Press, Boca Raton, 2013.
- E. Hollifield e A. Zarn. *A Survival Analysis of the Duration of Olympic Records* URL <https://arxiv.org/abs/1207.6133v1> [acedido em 01-06-2018], 2012
- O. G. Stevenson e B. J. Brewer. Bayesian survival analysis of batsmen in Test cricket. in *Journal of Quantitative Analysis in Sports*, 13, 1, pp. 25-36, 2017
- G. R. Monteiro, R. J. S. Silva. The length of stay in community physical activity program does not exceed two years. in *Revista Brasileira de Cineantropometria e Desempenho Humano*, 16, 6, 608-617, 2014
- E. L. Kaplan e P. Meier. Nonparametric Estimation from Incomplete Observations. in *Journal of the American Statistical Association*, 58, 282, pp. 457-481, 1958
- David Cox. Regression Models and Life-Tables. in *Journal of the Royal Statistical Society*, B34, 2, pp. 187-220, 1972
- J. D. Kalbfleisch e R. L. Prentice. *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, Nova Iorque, 2002
- David Cox e E. J. Snell. A general definition of residuals (with discussion). in *Journal of the Royal Statistical Society*, A, 30, pp. 248-275, 1968
- D. A. Schoenfeld. Partial residuals for the proportional hazards regression model. in *Biometrika*, 69, pp. 127-141, 1982
- D. Nevo e Y. Ritov. Around the goal: examining the effect of the first goal on the second goal in soccer using survival analysis methods. in *Journal of Quantitative Analysis in Sports*, 9, 2, pp. 165-177, 2013
- P. M. Grambsch e T. M. Therneau. Proportional hazard tests and diagnosis based on weighted residuals. in *Biometrika*, 81, pp. 515-526, 1994