

As Crônicas da Sobrevivência

Alisson Rosa

Resumo

A maioria das análises ¹ empregando o banco de dados utilizado nesse trabalho remove as observações que não tiveram o evento de interesse, isso evidentemente é um empreendimento equivocado, pois tais observações também carregam informações importantes, o que foi desenvolvido no presente trabalho portanto é uma análise sobrevivência do tempo de vida dos personagens dos livros As Crônicas de Gelo e Fogo escritos por George R. R. Martin. Foi avaliado as funções de sobrevivência para as as variáveis selecionadas, juntamente com um modelo de regressão de Cox para avaliar influência das covariáveis na função taxa de falha, toda análise foi desenvolvida usando a linguagem R [1] e o código elaborado pode ser consultado aqui.

Sumário

1	Introdução	2
2	Estimando Quantidades Básicas	2
2.1	Análise de Sobrevivência	2
2.2	Metodologia da criação do banco de dados	2
2.3	Análise Descritiva	2
3	Estimando Quantidades Básicas	5
3.1	Estimando a Sobrevivência	5
4	Escolha de um Modelo Paramétrico	9
4.1	Modelo de regressão de Cox	9
4.2	Ajuste do modelo	9
4.3	Análise de diagnóstico	10
4.4	Interpretação dos coeficientes	10
4.5	Revisitando gráficos importantes	11
5	Conclusão	11
6	Bibliografia	11

¹Podem ser consultadas aqui

1 Introdução

As Crônicas de Gelo e Fogo é uma série de livros de fantasia épica escrita pelo romancista e roteirista norte-americano George R. R. Martin. Martin começou a desenvolvê-la em 1991 e o primeiro volume foi lançado em 1996. Originalmente concebida para ser uma trilogia, a saga agora (2022) consiste em cinco volumes publicados, com mais dois planejados. Há também três contos derivados e algumas novelas que consistem de resumos dos romances principais.

Há três argumentos principais na história, que se tornam cada vez mais interligados: a crônica de uma guerra civil dinástica entre várias famílias concorrentes pelo controle dos Sete Reinos; a ameaça crescente das criaturas sobrenaturais, que habitam além de uma imensa muralha de gelo ao Norte; e a ambição de Daenerys Targaryen, a filha exilada de um rei assassinado em uma outra guerra civil treze anos antes, prestes a voltar à sua terra e reivindicar seu trono de direito.

Os livros de As Crônicas de Gelo e Fogo foram adaptados para diversos formatos, como jogos de videogame, histórias em quadrinhos, bonecos em miniatura e uma série de TV intitulada Game of Thrones. A atração televisiva apresentou a saga a um maior número de leitores e lhe trouxe maior notoriedade, fazendo com que os quatro primeiros volumes da série surgissem entre os dez primeiros colocados na referencial lista de mais vendidos do jornal norte-americano The New York Times em 2011.

Vamos nesse breve ensaio fazer uma análise de sobrevivência estatística dos personagens dos livros, onde o banco de dados aqui utilizado está disponível no [Kaggle](#).

2 Estimando Quantidades Básicas

2.1 Análise de Sobrevivência

A análise de sobrevivência difere dos métodos estatísticos clássicos pois temos dados censurados, que são observações a qual o evento de interesse por algum motivo não pode ser mensurado. Aqui, o nosso evento de interesse é o **tempo para morte** de um personagem, e se o personagem não faleceu, será classificado como censurado, e seu respectivo tempo de censura será o último livro que apareceu. Assim, nossa variável tempo (T) possui somente 5 valores possíveis, que são respectivamente a numeração dos livros.

2.2 Metodologia da criação do banco de dados

As crônicas de gelo e fogo apesar de possuir “apenas” 5 livros é uma obra extensa possuindo uma quantidade enorme de personagens, dessa maneira fica claro que o número de personagens que morre é muito menor que os que não morrem, portanto para coesão da análise foi removido todos os indivíduos que não estavam associados a alguma família relevante². O banco de dados em sua forma “original” não está pronto para a análise de sobrevivência, pois é necessário uma variável que informe se o tempo observado foi de censura ou do evento de interesse, assim essa variável foi construída para análise, onde o banco modificado pode ser encontrado [aqui](#).

2.3 Análise Descritiva

Como temos muitos personagens, também temos uma quantidade considerável de associações³, assim é interessante avaliarmos a quantidade de indivíduos dentro dessas associações, como também a quantidade de mortos respectivamente. O que pode ser avaliado pela Figura 1.

²Onde relevante é definido arbitrariamente pelo autor que vos escreve.

³No texto, associações na maior parte referem-se a famílias, porém existem também associações a quais não são famílias.

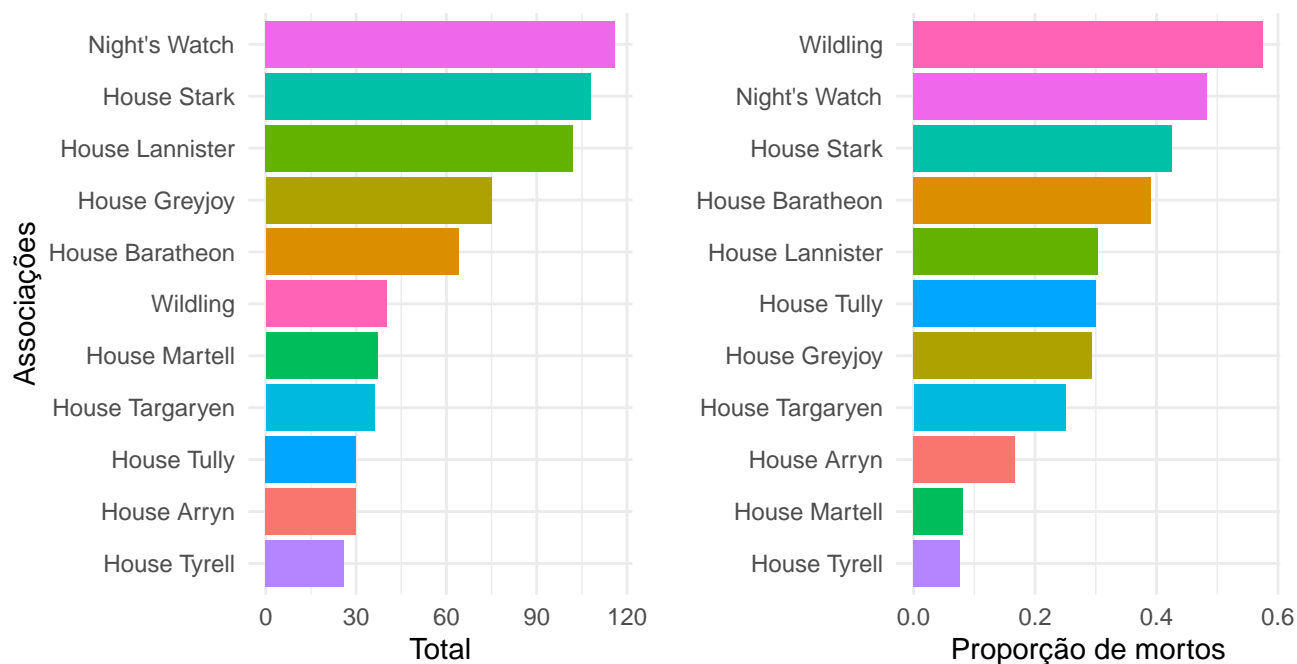


Figura 1: Comparação entre quantidade de membros nas associações e mortos

Notamos portanto que a proporção de personagens mortos não segue a proporção de quantidade de membros das associações, pois os Windlings (Povo Livre em português) é a associação que teve mais baixas e associação que mais tem membros é a Night's Watch (Patrulha da Noite em português). Note que três famílias se destacam tanto em quantidade de membros e em mortos, sendo elas os Starks, os Lannister e os Baratheon. Os Greyjoy também estão nas famílias com mais membros, estão em 4 lugar, entretanto tal posição não se mantém na quantidade de mortos.

Dado o que foi comentado, é interessante juntar algumas famílias que possuem poucos membros em um outro grupo para análise ser mais focada, assim vamos agrupar as casas Martell, Tully, Tyrell, Arryn em um grupo chamado de "outros", vale ressaltar que os Targaryen também possuem poucos membros, porém são peça chave da obra.

Pela Figura 2, podemos averiguar a proporção de personagens de cada sexo na obra, junto com a quantidade de mortos por sexo.

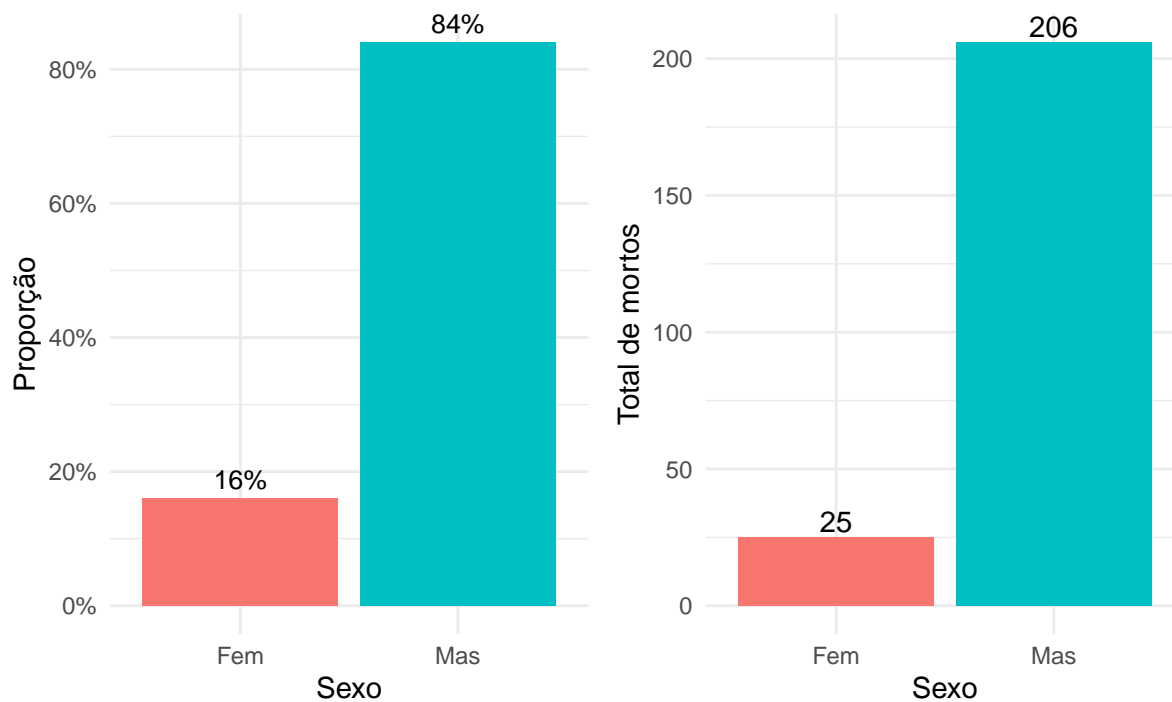


Figura 2: Proporção de personagens por sexo e quantidade de mortos

Notamos o predomínio na quantidade de personagens masculinos, o comportamento de mortos segue a proporção da quantidade. Pela 3 temos a quantidade de personagens por classe social, junto com a quantidade de mortos por classe.

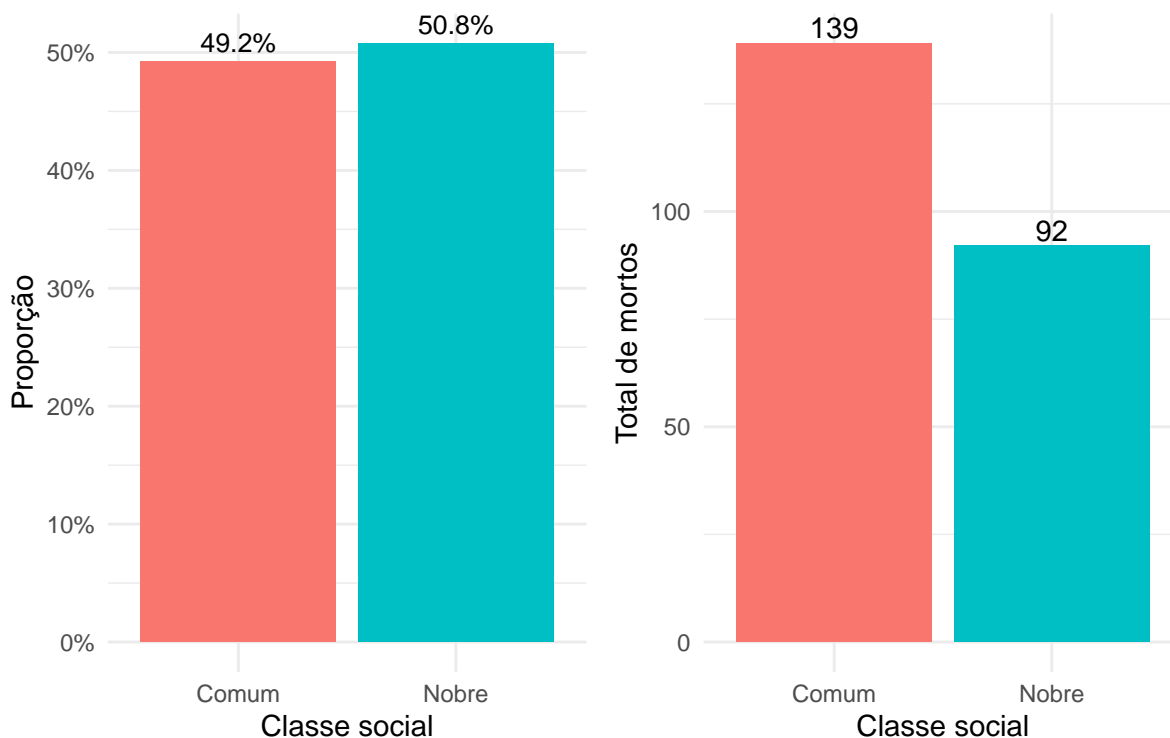


Figura 3: Proporção de personagens por classe social e quantidade de mortos

Diferente do que vimos na Figura 2, para a classe social a quantidade de mortos não segue a proporção de indivíduos na classe, pois temos mais personagens comuns morrendo.

3 Estimando Quantidades Básicas

A função de sobrevivência S nos informa qual a probabilidade de uma observação vir a ter o evento de interesse a partir do tempo t , em termos formais podemos definir como:

$$S(t) = P(T > t) = 1 - F(t) \quad (1)$$

desse modo podemos avaliar a função de sobrevivência estratificada por grupos, avaliando tanto em termos visuais e inferenciais a possibilidade de existir diferença em S . Uma quantidade também interessante e útil mais a frente é função taxa de falha acumulada Λ , que pode ser definida como:

$$\Lambda(t) = -\log(S(t)) \quad (2)$$

3.1 Estimando a Sobrevivência

Sabemos que a proporção de personagens do sexo feminino é de apenas aproximadamente 16%, desse modo uma pergunta a se fazer é se existe diferença na sobrevivência entre os sexos, o que pode ser avaliado pela Figura 4

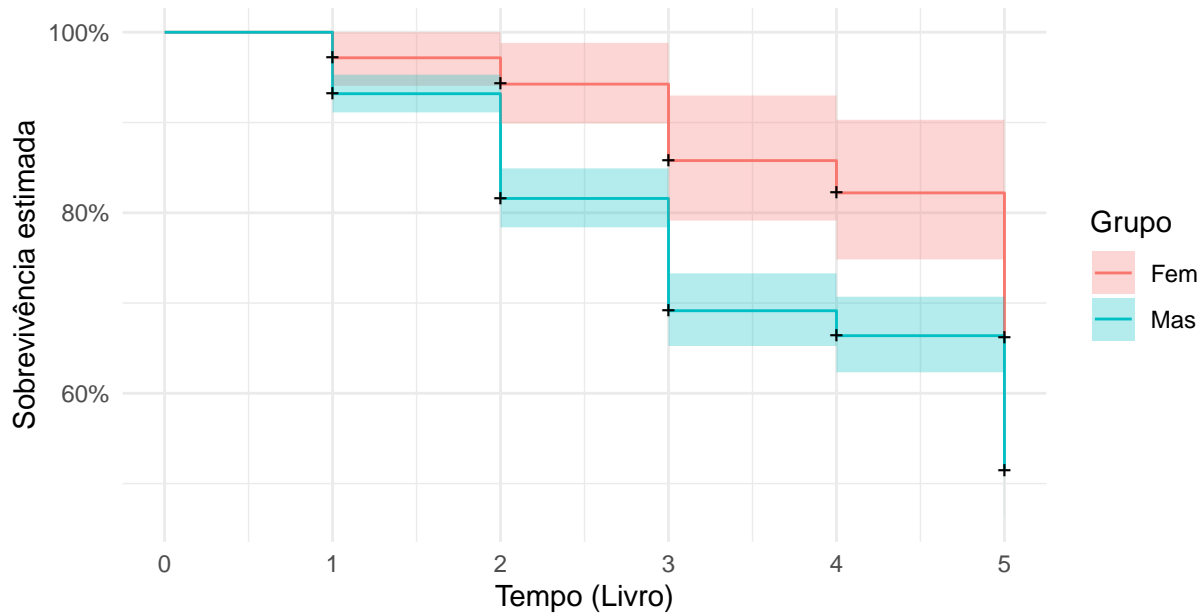


Figura 4: Função de sobrevivência estimada para ambos os sexos

Assim, a Figura 4 nos fornece indícios de que existe diferença no tempo de sobrevivência entre os sexos, pois a curva estimada de S para o sexo feminino está pontualmente superior que a do sexo masculino para todos os tempos, e mesmo em termos de intervalos de confiança existe pouca interseção.

Também é importante verificar se existe diferença nas funções de sobrevivência para a classe social do personagem, o que pode ser visto pela Figura 5.

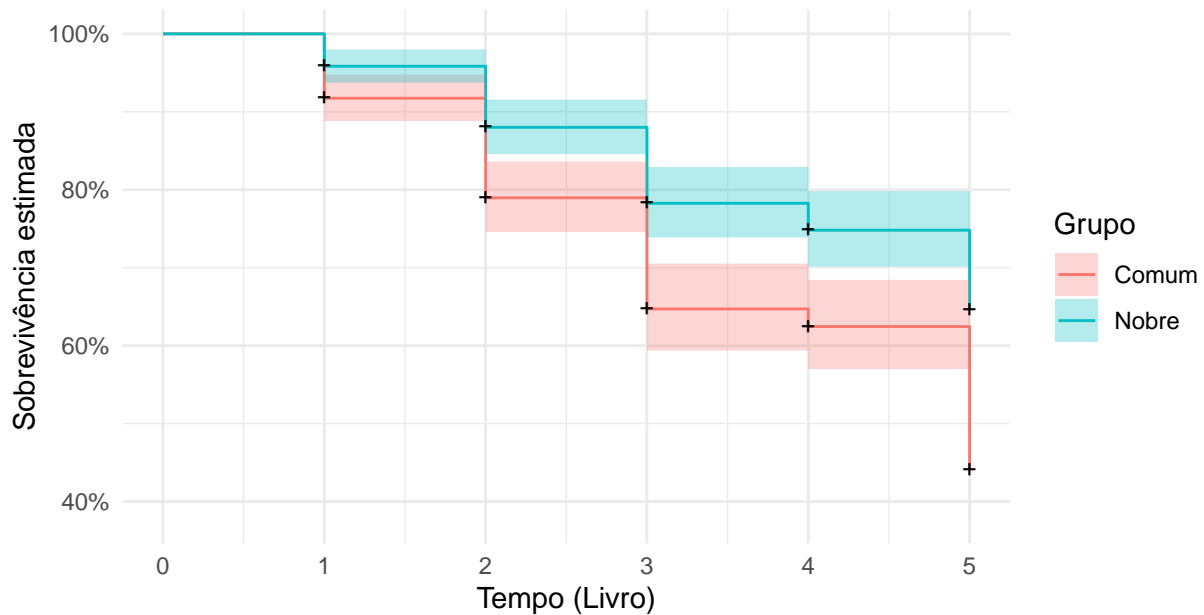


Figura 5: Função de sobrevivência estimada para as classes sociais

Como é de se esperar ⁴, personagens que são de uma família nobre tem uma sobrevivência estimada superior aos que

⁴Warui kizoku wo yattsukero da - Moriarty the Patriot

são de família comum.

Dado que as Figuras 4 e 5 indicaram diferença no tempo de sobrevivência para sexo e nobreza, é importante se perguntar se a interação entre esses grupos mantém tal diferença, o que pode ser avaliado pela Figura 6.

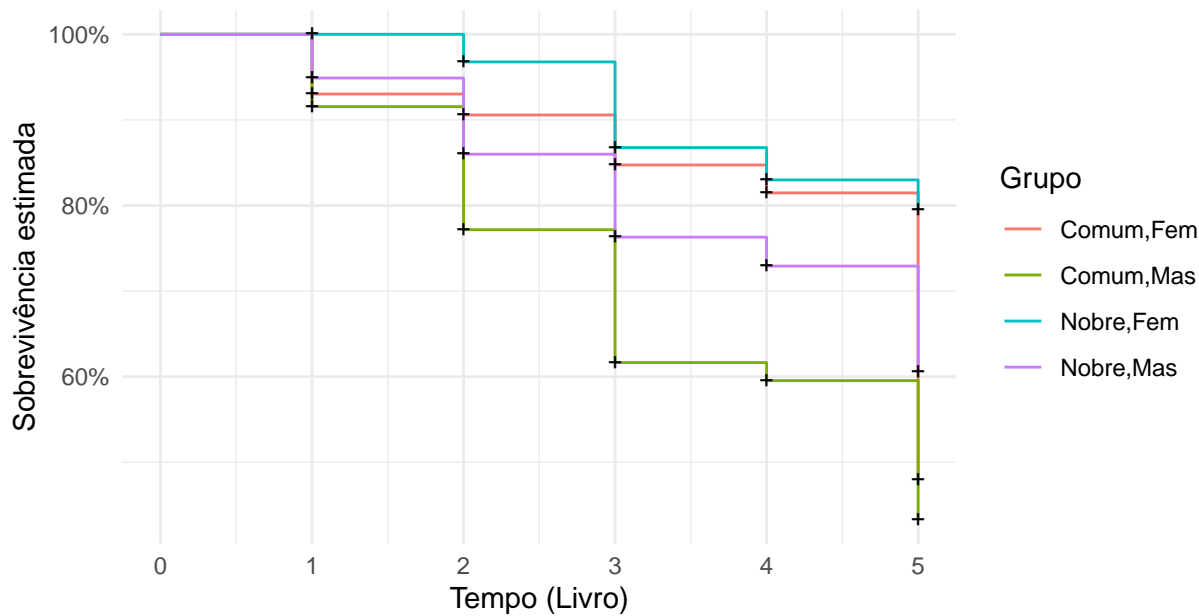


Figura 6: Função de sobrevivência estimada para as classes sociais e sexos

Constata-se portanto que o comportamento individual das variáveis se mantém nas interações, pois para cada sexo e classe social a sobrevivência com valores superiores é sempre a da Nobreza, nesse mesmo sentido até mesmo o sexo feminino da classe comum possui maior taxa de sobrevivência que a do sexo masculino nobre ⁵.

Como vimos na seção anterior os Starks tem uma grande quantidade de membros e também uma grande quantidade de mortos, portanto é pertinente analisar a função de sobrevivência para eles e também com algumas famílias, o que pode ser visto pela Figura 7.

⁵Contexto **importa**: A Figura 6 sem contexto pode facilmente tornar-se “como mentir com estatística”.

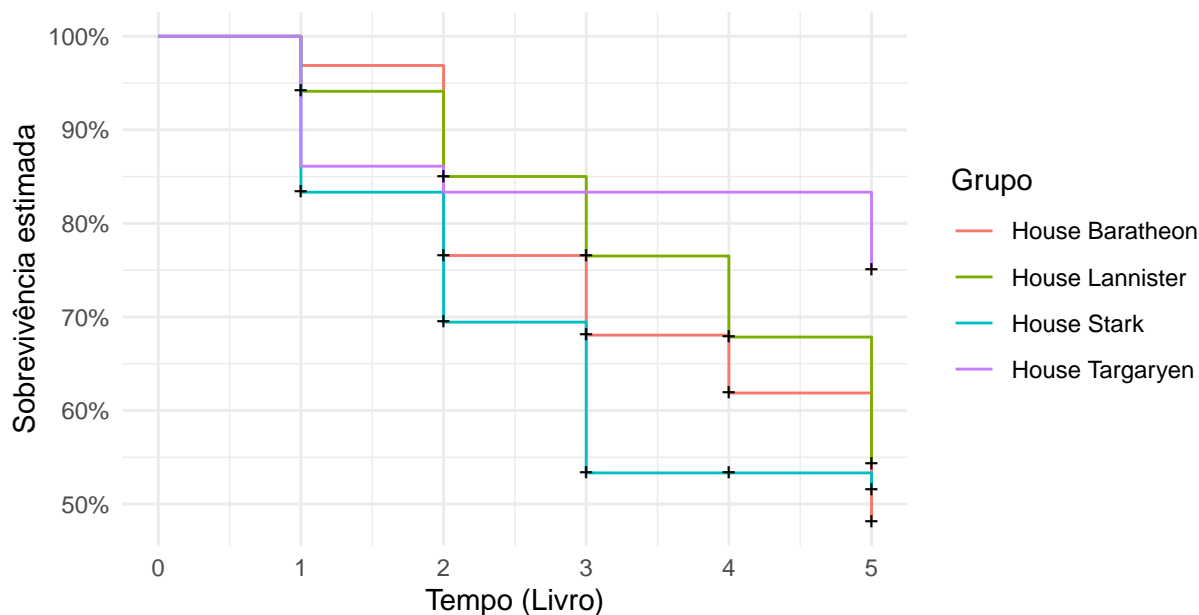


Figura 7: Função de sobrevivência estimada para as Associações 1

Dessa forma a Figura 7 nos indica que a vida de um Stark não é fácil, a cada livro lançado até o livro 3, a sobrevivência para um Stark despenca, e no livro 3 ⁶ podemos ver a maior queda. Podemos notar também que os Baratheon iniciam com uma sobrevivência superior ao Lannister, mas é um comportamento que não se mantém a partir do livro 2.

Podemos além disso avaliar a sobrevivência para o Povo Livre, a Patrulha da Noite e a união das famílias que agora são denotadas por “outros”, o que pode ser observado pela Figura 8.

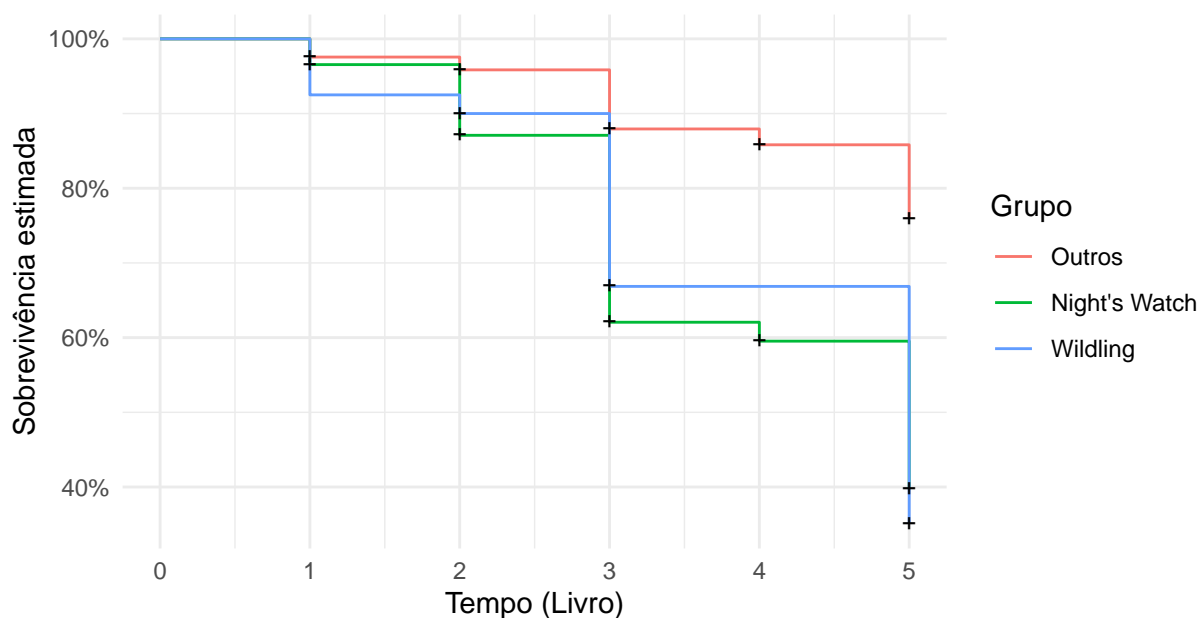


Figura 8: Função de sobrevivência estimada para as Associações 2

Notamos que até o livro 2 temos um comportamento semelhante, porém o Povo Livre e a Patrulha da Noite tomam

⁶Red wedding para os mais familiarizados.

um grande abate no livro 3 ⁷, e como era de se esperar, no livro 5 o grupo dos “outros” formado somente por nobres mantém a maior sobrevivência.

Faz-se então importante realizarmos um teste de hipóteses para avaliar as decorrências anteriores, aqui vamos utilizar o teste log-rank, onde em termos gerais possui a seguinte hipótese.

$$H_0 : \text{As funções de sobrevivências dos grupos são iguais.}$$

O p-valor do teste para cada grupo é indicado pela Tabela 1.

Tabela 1: P-valores fornecidos pelo teste log-rank

Variável	P.valor
Sexo	<0.01
Nobreza	<0.01
Associações 1	0.1
Associações 2	<0.01

Desta forma podemos averiguar que o único grupo a qual o teste não captou diferença na função de sobrevivência foi o de Associações 1. que é formado pelos Starks, Lannister, Targaryen e Baratheon. Para Sexo, Nobreza e Associações 2, existem evidências de diferença nas funções de sobrevivência.

As análises anteriores são pertinentes para a seleção das covariáveis do modelo de regressão, pois como vimos existem evidências de diferença nas funções de sobrevivência nesse grupos, em outras palavras dependendo do grupo a qual o personagem está situado seu tempo de sobrevivência pode ser diferente.

4 Escolha de um Modelo Paramétrico

4.1 Modelo de regressão de Cox

Vamos nessa seção ajustar um modelo de regressão de Cox a qual foi introduzido pela primeira vez em [2]. Onde vamos utilizar como covariáveis, o sexo do personagem e se é da nobreza, porém antes de ajustar o modelo precisamos verificar se o pressuposto de taxa de falha proporcionais não foi quebrado. Dessa maneira devemos primeiro realizar um teste para verificar se existem evidências de não proporcionalidade, em termos gerais o teste possui a seguinte hipótese:

$$H_0 : \text{Os riscos são proporcionais}$$

Realizando o teste temos como p-valor associado 0.067, 0.879 para as variáveis Sexo, Nobreza, respectivamente. Portanto indicando que não existem indícios de falha no pressuposto.

Vamos agora então ajustar o modelo e fazer algumas análise de diagnóstico

4.2 Ajuste do modelo

O modelo de Regressão de Cox depende de parâmetros $\beta's$ desconhecidos e nessa abordagem constantes, que medem o efeito das covariáveis sobre a função taxa de falha, entretanto na prática possuímos apenas algumas observações ao nosso alcance, nesse caso, somente os personagens a qual o autor citou nos livros, mas existem outros no universo a qual o nome não foi mencionado, assim, devemos utilizar algum método de estimação para encontrar uma aproximação para os $\beta's$ para que o modelo possa ser ajustado.

Ajustando o modelo com os dados observados, temos a Tabela 2 com o resumo do ajuste.

Tabela 2: Resumo do ajuste

	Coefficientes	exp(Coefficientes)	Erro padrão	P.valor
GenderMas	0.546	1.73	0.212	0.01
NobilityNobre	-0.510	0.60	0.135	<0.01

⁷Para os íntimos: Exploração além das muralhas.

Como esperado, as covariáveis são significativas para explicar a taxa de falha, notamos que se o personagem é do sexo masculino, ele tem uma aceleração na taxa de falha pois a exponencial de seu coeficiente estimado é maior do que 1 o que não acontece para um personagem nobre, dado a exponencial do coeficiente estimado é 0.6 indicando desaceleração na taxa de falha.

4.3 Análise de diagnóstico

O modelo de regressão de Cox não se adequa a qualquer situação, assim como qualquer modelo estatístico clássico devemos fazer uma avaliação geral da adequação. Aqui vamos utilizar o Resíduo de Cox-Snell (e) a qual foi definido em [3], tal resíduo se o modelo estiver corretamente ajustado pode ser visto como uma amostra censurada de uma distribuição exponencial com parâmetro igual a 1, assim o gráfico de e_i versus $\Lambda(e_i)$, vai ser uma reta com inclinação 1, pois dada a definição de Λ em (2), e sabendo que a função acumulada de uma variável aleatória T com distribuição exponencial com parâmetro igual a 1 é dada por $\exp(-t)$, por consequência temos:

$$\Lambda(e_i) = \Lambda(S(e_i)) = -\log(\exp(-e_i)) = e_i \quad (3)$$

Dessa maneira, podemos analisar tal comportamento pela Figura 9.

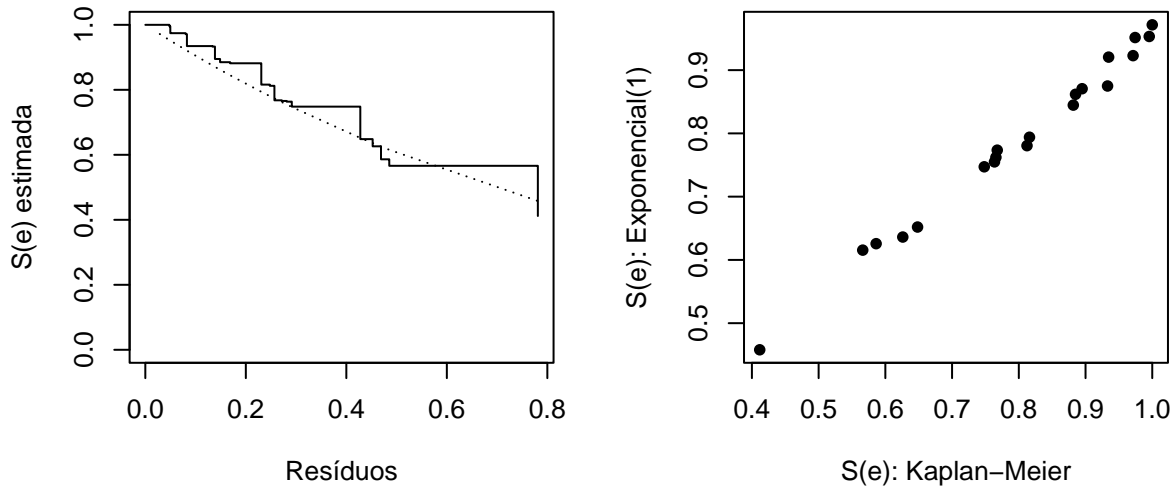


Figura 9: Resíduos de Cox-Snell

Podemos notar em ambas as Figuras que visualmente não existem evidências de não adequação do ajuste, pois na Figura á esquerda temos ambas as curvas bastante semelhantes enquanto na Figura á direita dispomos de um comportamento que se assemelha a bissetriz dos quadrantes ímpares.

4.4 Interpretação dos coeficientes

Uma grande vantagem que o modelo de Regressão de Cox dispõe é a interpretabilidade, pois como indicado em [4] se tivermos taxas de falhas proporcionais, conseguimos verificar que o efeito das covariáveis é acelerar ou desacelerar a taxa de falha.

Nesse caso um valor de 1 para a covariável **GenderMas** indica que o personagem é do sexo masculino, assim dado seu coeficiente estimado, temos que o risco de falha para um personagem masculino é 1.727 maior que para um personagem feminino, um número que formaliza em termos de taxa de falha o que foi visto na Figura 4 pela função de sobrevivência.

Na situação da covariável `NobilityNobre` que indica se o personagem é um nobre, temos um efeito de desaceleração da taxa de falha, pois seu coeficiente associado é menor do que 0, outra fato esperado dado o o comportamento da função de sobrevivência na Figura 5.

4.5 Revisitando gráficos importantes

Vamos aqui revisar algumas quantidades anteriores, porém agora sobre a estrutura do modelo de Regressão de Cox a figura 10 nos fornece a função de sobrevivência e o risco acumulado.

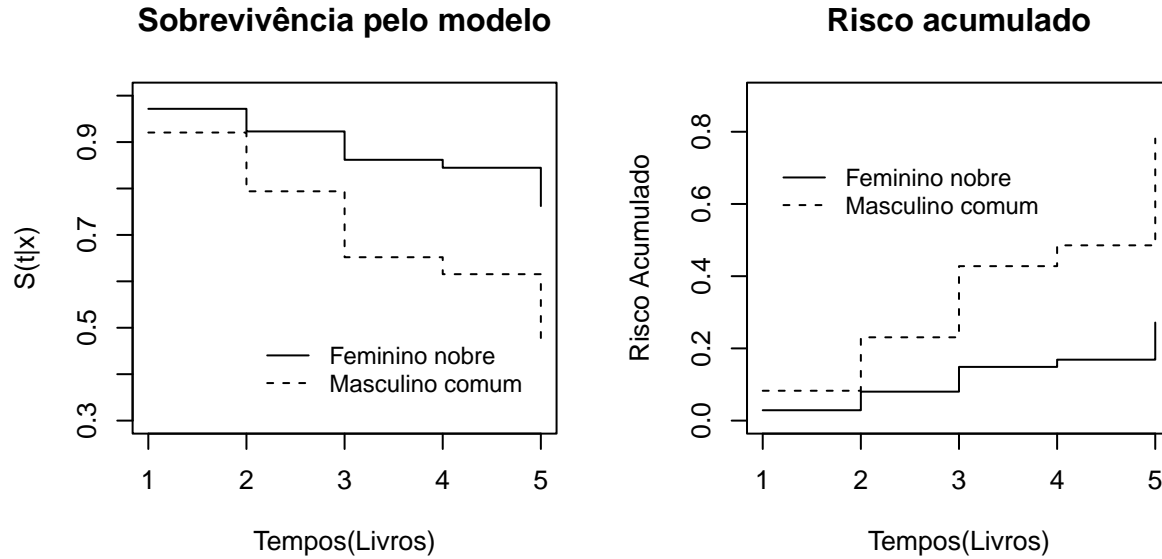


Figura 10: Função de Sobrevivência e Risco acumulado utilizando o modelo

Notamos assim que o modelo conseguiu assimilar os atributos que vimos ao longo da análise, pois um personagem feminino nobre possui uma sobrevivência maior e risco menor do que um masculino comum ao longo dos livros.

5 Conclusão

Contemplamos a fundo o comportamento da personagens na obra *As Crônicas de Gelo e Fogo*, onde notamos que a maioria dos personagens é do sexo masculino, e também os que mais morrem são masculinos, dentro das classes sociais os personagens que mais morrem não são nobres, apesar da quantidade de nobres e personagens de classe comuns na obra ser bastante similar. A associação que teve maior desfalque na função de sobrevivência foi os Starks, entretanto não foi a associação que mais teve mortes, que foi o Povo Livre. Notamos tanto visualmente quanto inferencialmente que os personagens femininos possuem uma função de sobrevivência superior que as dos masculinos, e a discrepância torna-se ainda maior quando coloca-se personagens femininos nobres.

Presenciamos que um modelo de Regressão de Cox utilizando como covariáveis se o personagem é do sexo masculino e se é nobre se ajustou corretamente, pois não temos evidências de falha no pressuposto de taxa de falha proporcional, assim como análise visuais não forneceram indícios de não adequação. Outra forte evidência de boa adequação é as funções de sobrevivência e risco acumulado estimadas, pois todas as características observadas no estudo foram captadas.

6 Bibliografia

- [1] R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2022. Available from: <https://www.R-project.org/>.

- [2] Cox DR. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1972;34:187–202.
- [3] Cox DR, Snell EJ. A general definition of residuals. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1968;30:248–265.
- [4] Colosimo SR Enrico Antonio e Giolo. *Análise de sobrevivência aplicada*. Editora Blucher; 2006.
- [5] Kleinbaum DG, Klein M, et al. *Survival analysis: A self-learning text*. Springer; 2012.
- [6] Huff D. *How to lie with statistics*. WW Norton & Company; 1993.
- [7] Wikipedia. A Song of Ice and Fire — Wikipedia, the free encyclopedia. <http://pt.wikipedia.org/w/index.php?title=A%20Song%20of%20Ice%20and%20Fire&oldid=64146846>; 2022.