

Estudo do tempo de conclusão dos cursos de graduação no CT-UFSM

Vítor Pereira

Resumo

Nesse presente trabalho para a disciplina de Análise de Sobrevida do Curso Estatística da UFSM, buscamos estudar e entender como variáveis sociais, econômicas e culturais afetam o tempo de formação dos alunos do CT, também da UFSM. Visando encontrar quais características tendem a aumentar ou diminuir o tempo de permanência até a formação dos estudantes, para isso utilizaremos técnicas para dados censurados, para captarmos informações dadas por dados incompletos de uma maneira sistemática e conhecida, na análise a censura é dada, por aqueles alunos que ainda estão realizando os cursos de graduação e para adequarmos essas observações para nossa análise utilizaremos os modelos de Tempo de Vida Acelerado ou Regressão de Cox.

Sumário

1	Introdução	1
2	Conhecendo os dados	2
2.1	Descrição das Variáveis	3
2.2	Visualização dos Dados	3
3	Princípio da Análise de Sobrevida	7
3.1	Teste de Logrank	10
4	Seleção do modelo	11
4.1	Seleção de covariáveis	12
5	Análise de Resíduo	14
6	Interpretação	17
7	Conclusão	19
	Apêndice - Código R	19

1 Introdução

Esse estudo é fruto de uma parceria entre o curso de Bacharelado em Estatística e a Pró-Reitoria de Planejamento (PROPLAN) da UFSM, sendo parte avaliativa da disciplina de Análise de Sobrevida e parte integrante da bolsa na Coordenadoria de Planejamento Informacional (COPLIN) buscando estudar os efeitos de covariáveis presentes no banco de dados de 2010 a 2022 da UFSM no tempo para a formação de alunos do Centro de Tecnologia (CT).

Para isso pretendemos empregar técnicas utilizadas na análise de sobrevida como Modelos de Tempo de Vida Acelerado e Regressão de Cox que possibilitam integrar covariáveis para modelar o tempo para a graduação do aluno, assim conseguimos integrar covariáveis que tragam informações sociais e econômicas e a possibilidade de incorporar dados censurados, assim incluindo a informação fornecida por dados incompletos, i.e., observações que ainda não alcançaram o evento de interesse, no caso do presente trabalho a conclusão do curso, mas temos a informação que o

aluno demorará, pelo menos o tempo que está para se formar, assim alunos em situação de permanência são os nossos dados censurados nesse estudo.

Desse modo, estamos desconsiderando as observações de alunos que evadiram dos cursos de graduação, para considerá-los precisaríamos de outra abordagem, a de riscos competitivos, simplificando, são situações em que as causas de falhas competem entre si, pois não podem ocorrer juntas, ou seja, se um aluno evadir de um curso de graduação no CT, nunca desenvolverá o evento de interesse, que é a formação. Em estudos clínicos, riscos competitivos, também são chamados de fração de cura, onde, por exemplo, buscamos estudar a recidiva do câncer, mas o paciente se cura.

Assim, esse estudo busca sugerir com técnicas estatísticas fatores que diminuem ou aumentam o tempo para a formação, visando informar sobre essas características dos indivíduos em estudo, aos responsáveis, para receberem amparo e apoio para terem a formação no tempo estabelecido, igualmente, buscamos comparar com outros estudos desenvolvidos na disciplina e em outras referências teóricas presente.

2 Conhecendo os dados

Toda boa análise estatística, deve-se começar de sua parte mais básica e fundamental, os dados, por isso realiza-se nessa seção uma análise descritiva simples, mas informativa das variáveis presentes na nossa base de dados, na Tabela 1 podemos ter um vislumbre das observações iniciais do banco

Tabela 1: Observações Iniciais do Banco

INICIO	SEXO	ETNIA	INGRESSO	COTA	CHAMADA	CURSO	DURACAO	CENSURA	TEMPO
2021	2	6	SiSU	Universal	Chamada	Engenharia Civil	5	0	1
2021	2	1	SiSU	Universal	Chamada	Engenharia Civil	5	0	1
2021	1	1	SiSU	Racial	Chamada	Engenharia Civil	5	0	1
2021	2	1	SiSU	Racial	Chamada	Engenharia Civil	5	0	1
2021	2	6	SiSU	Racial	Chamada	Engenharia Civil	5	0	1

^a Algumas variáveis sofreram alterações no nome para melhorar visualização da tabela

No entanto, além de visualizar as principais variáveis do banco, também devemos construir algumas métricas para resumir e verificar sua distribuição e concentração, assim a Tabela 2 nos traz informações quanto as variáveis “numéricas”.

Tabela 2: Descrição das Variáveis Não-categóricas

variable	mean	median	sd	min	max	na_count
ANO_INGRESSO	2015.920	2016	3.528	2010	2022	0
ANO_EVASAO	2017.351	2018	2.415	2010	2022	3497
ID_SEXO	1.294	1	0.456	1	2	0
ID_ETNIA	1.653	1	1.424	1	6	0
DURACAO	4.872	5	0.334	4	5	0
CENSURA	0.249	0	0.432	0	1	0
TEMPO	3.417	3	2.263	0	12	0

^a Variáveis inicialmente tratadas como numérica pelo SoftWare R

A Tabela 2 possui informações valiosas, por exemplo temos 3203 observações censuradas, pois temos esse número de NA's na variável ANO_EVASAO, ID_SEXO e ID_ETNIA estão sendo tratados como variáveis numéricas, quando são categorias, assim precisamos mudar sua tipificação e analisar as variáveis categóricas.

2.1 Descrição das Variáveis

Para melhorar a compreensão dos trabalhos, temos uma breve da descrição de cada variável e suas categorias (se houverem):

- **ANO_INGRESSO**: Ano de ingresso no curso de graduação do CT, defini o tempo de início da variável tempo, o objetivo de interesse no estudo.
- **ANO_EVASAO**: Ano de saída do curso, defini o fim da variável tempo ou caso esteja incompleta, a sua censura, caso não exista conclusão ou permanência no curso, a observação será desconsiderada.
- **ID_SEXO**: Feminino ou Masculino.
- **ID_ETNIA**: Sem informação, Branca, Parda, Preta, Indígena ou Amarela.
- **INGRESSO**: Forma de Ingresso na graduação: Transferência, SiSU, Vestibular, Convênios, Processo Seletivo Seriado, Reingresso, Mobilidade Acadêmica, Refugiados, Seleção ou Outros.
- **COTA**: Cota utilizada para o ingresso no curso, com toda cota social, racial ou PCD, também sendo de Escola Pública, assim tem-se: Sem informação, Universal, Racial, Social, PCD, Escola Pública, Social e Racial, Racial e PCD, Social e PCD ou Social, Racial e PCD.
- **CHAMADA**: Se o aluno entrou pela primeira chamada (Listão) ou Chamadas Orais (Chamada): Sem informação, Listão ou Chamada.
- **SITUACAO** Situação do aluno quanto a conclusão/evasão no curso: Permanência, Desistência, Conclusão ou Falecido.
- **NOME_CURSO_AJUSTADO** Nomes de todos os cursos de graduação presentes no CT: Engenharia Civil, Engenharia Acústica, Engenharia Elétrica/CT, Engenharia Mecânica/CT, Sistemas de Informação/CT, Engenharia Aeroespacial, Ciência da Computação, Engenharia de Produção, Engenharia de Controle e Automação, Engenharia Química, Engenharia Sanitária e Ambiental/CT, Engenharia de Computação, Arquitetura e Urbanismo/CT ou Engenharia de Telecomunicações.
- **TURNO_CURSO** Todos os cursos do CT presentes no banco são Diurnos.
- **DURACAO** Os cursos tem duração de 4 ou 5 anos.
- **TEMPO** Tempo até o evento de interesse ou censura, varia de 0 a 12.
- **CENSURA** Verifica se o indivíduo está realizando a formação ou concluiu, para podermos adicionar as informações dos dados censurados.

2.2 Visualização dos Dados

Para ampliar significativamente nossa compreensão sobre os nossos dados, devemos fazer a análise mais simples, mas muito informativa, a análise gráfica, nessa seção vamos apenas apresentar gráficos com a quantidade de cada covariável e analisar as características mais relevantes, nessa parte inicial ainda não serão retirados os estudantes desistentes.

Para a Figura 1, percebe-se uma característica, inata do corte de tempo que estamos analisando, como o banco que utilizamos é de 2010 a 2022, tem-se que a evasão, aumenta em forma de escada até 2019, o que deve-se pelos alunos de 2010 e 2011, começarem a se formar por 2015 e 2016, no entanto, os que não concluíram o curso nesse período e não desistiram, foram concluído o curso posteriormente antes de jubilar, assim como pelo maior ingresso de alunos após 2014, o que aumenta tanto a desistência, quanto conclusão. Em 2022, há poucos ingressos, pois só existem dados do SiSU do primeiro semestre (2022/1).



Figura 1: Proporção do ano de ingresso e evasão dos Alunos do CT

Temos a confirmação do estereótipo dos alunos do CT, com a Figura 2, sendo composto em sua enorme maioria, por alunos homens e brancos. Pardos, Indígenas e Pretos estão em menor número que mulheres historicamente, apesar das ações afirmativas.

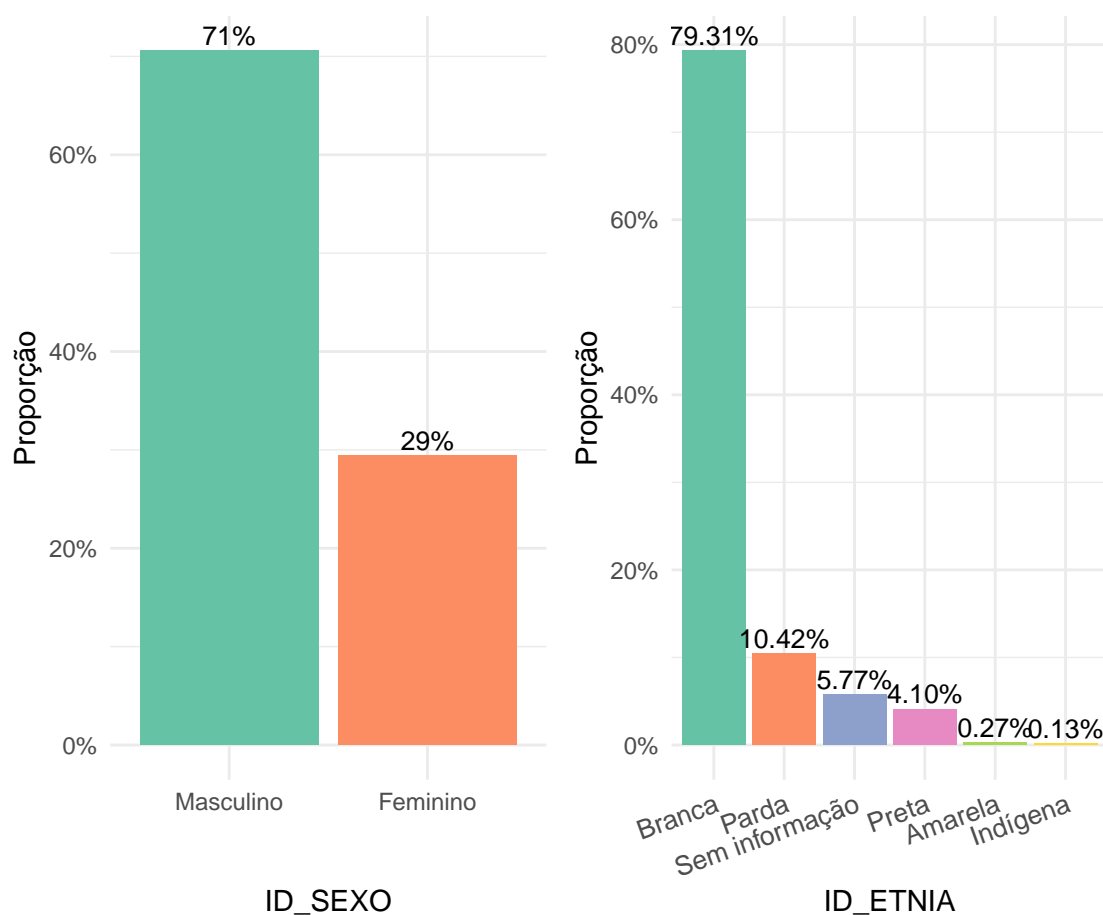


Figura 2: Proporção do Sexo e Etnia dos Alunos do CT de 2010 a 2022

Em relação a forma de ingresso e cota, temos que SiSU e Vestibular são realmente as formas preponderantes de entrada na faculdade, com o SiSU mesmo obrigatório a menos anos que o Vestibular no banco de dados, já possuindo a maior parte dos alunos, Processo Seletivo Seriado e Transferências também ocupam parcela importante da forma de ingresso, de acordo com a Figura 3. Para as cotas, temos a cota Universal com percentual maior que 50%, seguido pela cota Racial, Social, Escola Pública e Sem informação próximos aos 10%, as cotas para PCD não somam 1%.

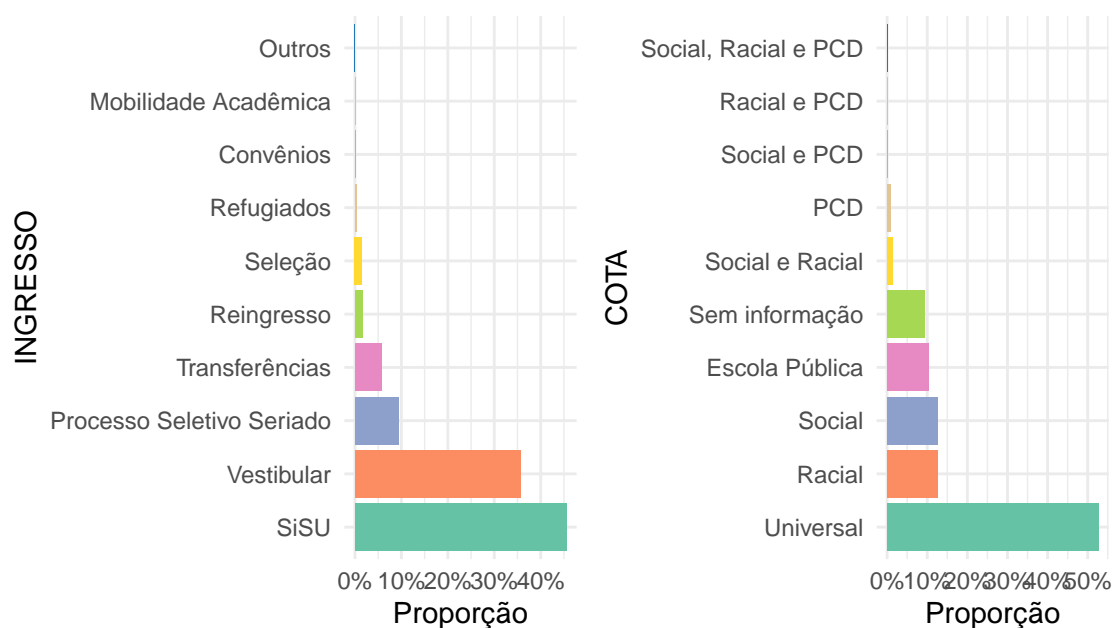


Figura 3: Proporção da forma de ingresso e cota dos Alunos do CT de 2010 a 2022

Para a descrição técnica dos cursos, temos que a maior parte dos alunos entra pelo Listão e 5% não tem informação, quanto a duração dos cursos, tem-se que 87% dura 5 anos, conforme a Figura 4.

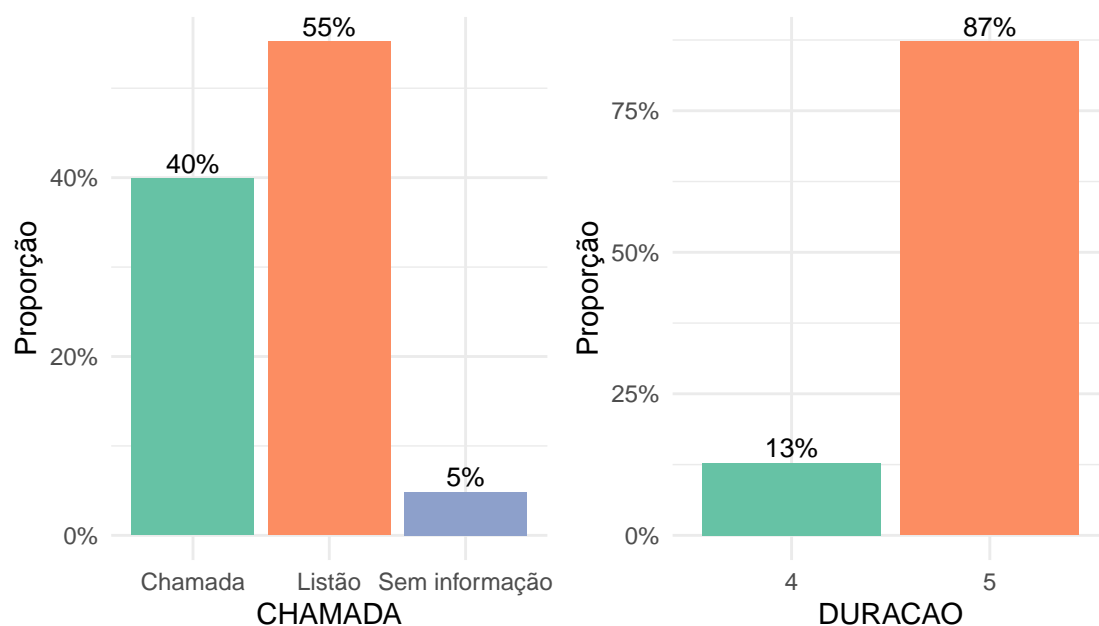


Figura 4: Proporção da chamada de entrada dos Alunos do CT de 2010 a 2022 e proporção de Duração dos cursos do CT

Nota-se que a distribuição dos alunos no cursos do CT são é simétrica, com grande parte ficando nas engenharias originais: Engenharia Civil, Química, Elétrica e Mecânica. A Figura 5, também demonstra que Engenharia de Produção (curso jovem, 2015), está entre os mais populares, assim como Engenharia Aeroespacial (2018), possui um número relevante de alunos.

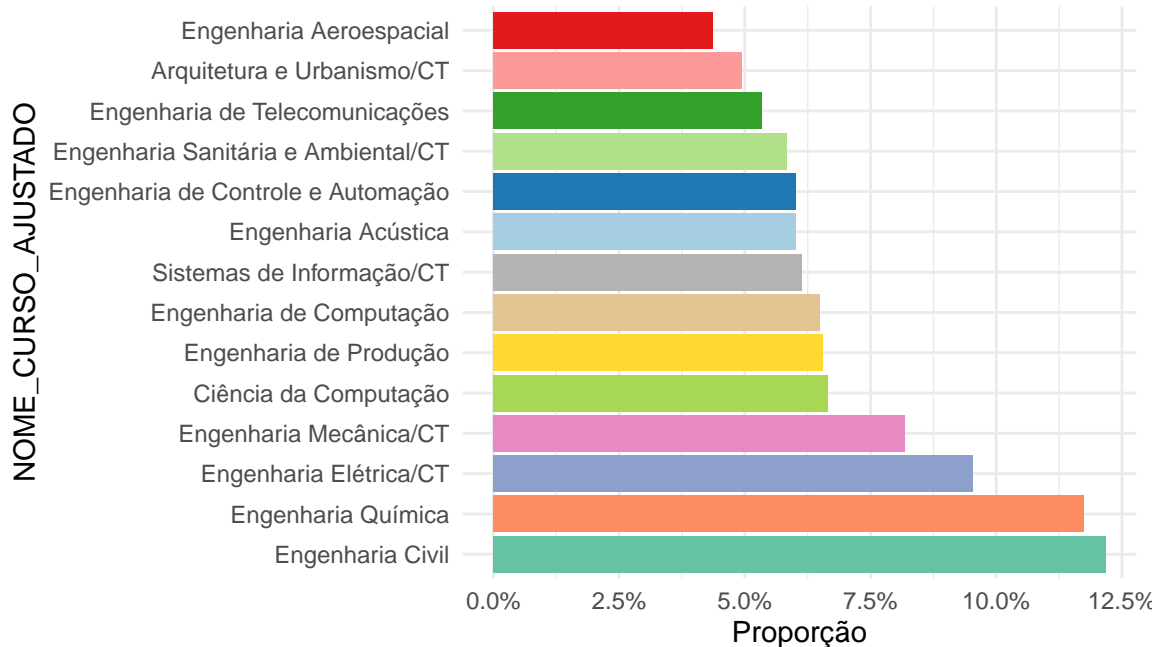


Figura 5: Proporção dos Alunos do CT de 2010 a 2022 em cada curso de graduação disponível do Centro

Para estudar de forma coerente o tempo para conclusão como um tempo de sobrevivência, vamos remover as censuras no tempo 0 e os alunos que efetivamente evadiram da faculdade.

3 Princípio da Análise de Sobrevivência

Precisa-se inicialmente conceituar e fundamentar os elementos primordiais da análise de sobrevivência, o tempo de falha utilizado nesse trabalho é “Tempo até a conclusão da graduação dos alunos do CT”, assim temos que a a função de sobrevivência é a probabilidade do discente não se formar até o tempo t . O tempo t é definido como $ANO_EVASAO - ANO_INGRESSO$, caso ANO_EVASAO seja NA , temos um dado censurado, e é substituído, pelo ano atual, 2022. Relembrando, que retiramos que as observações que evadiram efetivamente, pois não tem como atingir o tempo de falha, i.e., se o aluno evadir, não irá se formar.

Tem-se que uma das propriedades da função de sobrevivência é convergir para 0 com o tempo tendendo para infinito, assim como é observado na Figura 6, o estimador de Kaplan-Meier, não tem necessariamente essa propriedade, mas nota-se a tendência decrescente, assim pode-se se dizer que conforme o tempo aumento, diminui a probabilidade do discente não se formar.

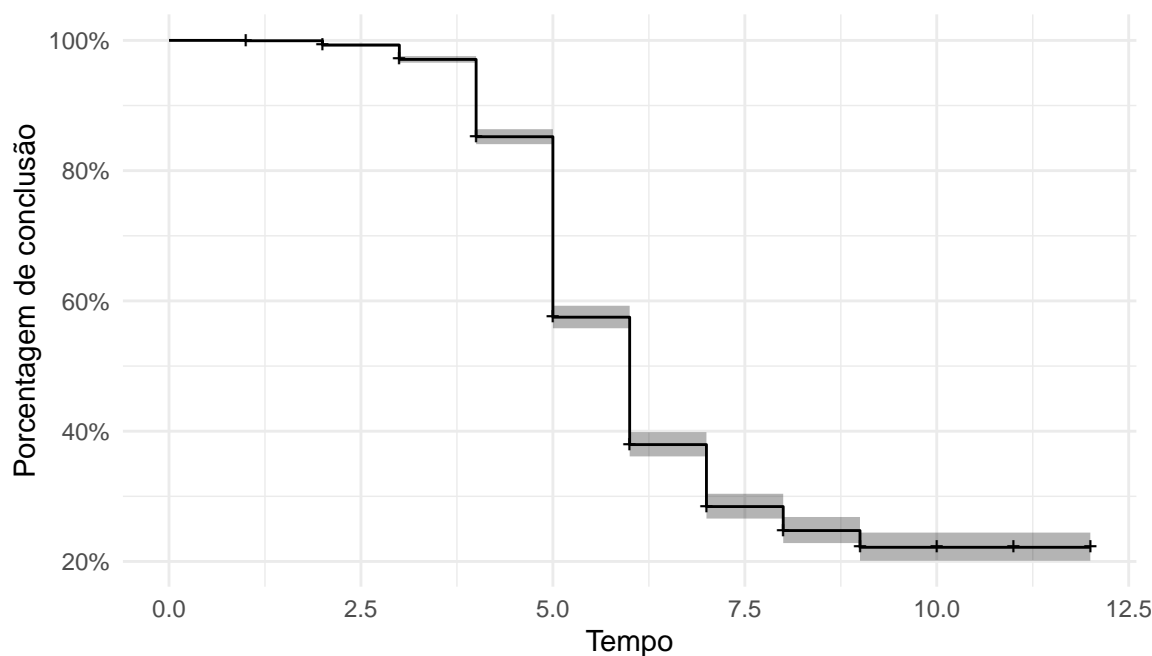


Figura 6: Função de Sobrevivência dada pelo estimador de Kaplan-Meier sem considerar covariáveis

Considerando Sexo ou Chamada como covariáveis, temos que para o Sexo Feminino, conforme o tempo aumenta, a diferença para o sexo Masculino também aumenta, assim a probabilidade do acadêmico não se formar também diminui. A Figura 7 mostra, que isso acontece para a variável Chamada, também, o tempo acentua a diferença entre as categorias.

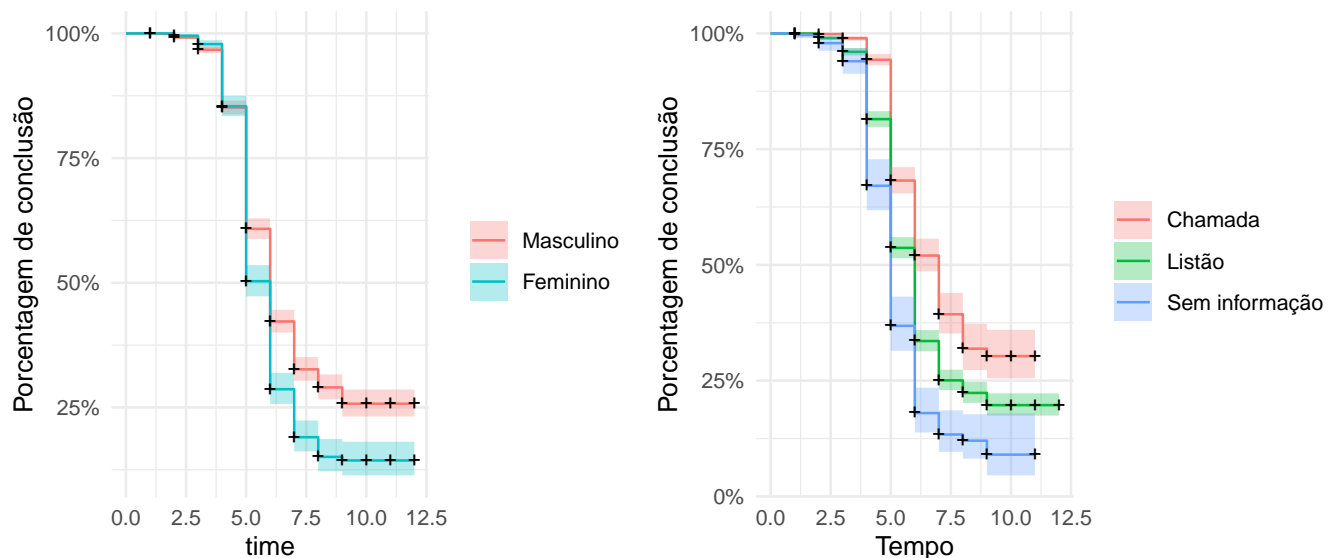


Figura 7: Função de Sobrevivência dada pelo estimador de Kaplan-Meier com a covariável Sexo (à esquerda) e Chamada (à direita)

Analisando a covariável Etnia, confirmamos o que intuitiva é pensado a curva para não formação que fica mais

abaixo é para a Etnia branca, seguido por Sem Informação e posteriormente, pela etnia Parda e Preta, com Indígenas e Amarelos não tendo observações suficientes para a curva decrescer, como demonstra a Figura 8.

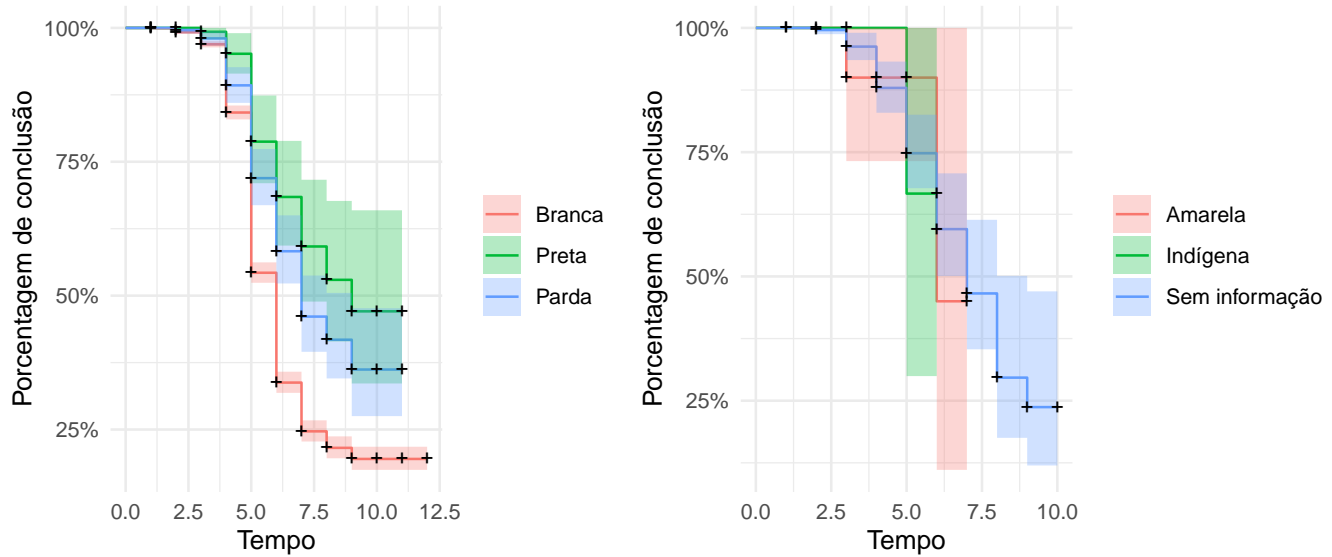


Figura 8: Função de Sobrevivência dada pelo estimador de Kaplan-Meier para a covariável Etnia

Na Figura 9, temos que as formas de ingresso que possuem menores sobrevivência de não formação são: Convênios, Processo Seletivo Seriado (PEIES), Vestibular e Transferências, respectivamente. Como o SiSU é uma forma recente, temos uma curva que ainda não atinge valores próximos a 0.

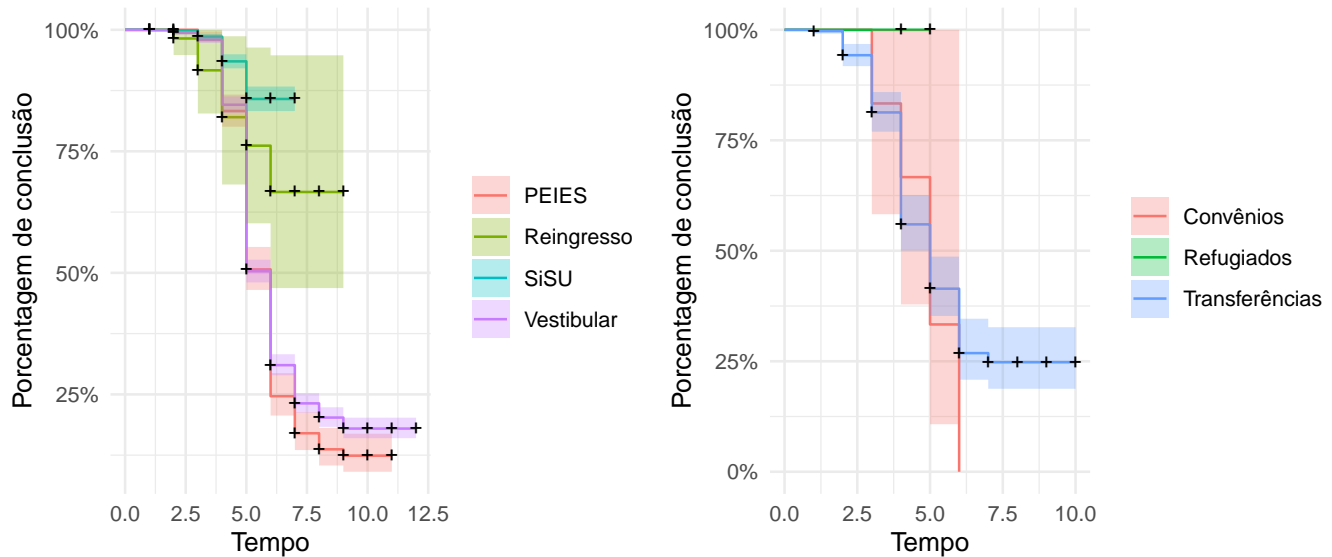


Figura 9: Função de Sobrevivência dada pelo estimador de Kaplan-Meier para a covariável Ingresso

Para a variável Cota, também confirmamos o pensamento imediato, as menores curvas são para as Cotas Universais

e de Escola Pública, com a categoria Sem Informação sendo a mais próxima da curva das anteriores, as observações com Cota PCD representam menos de 1 dos dados e não foi possível nem ajustar a curva, conforme podemos notar na Figura 10.

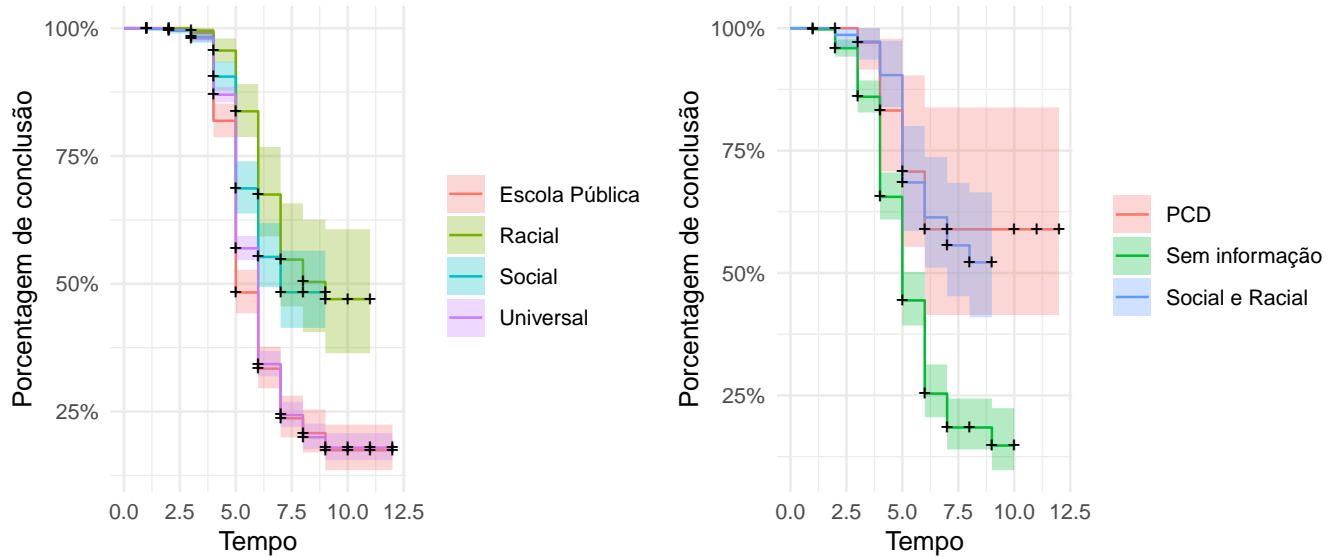


Figura 10: Função de Sobrevivência dada pelo estimador de Kaplan-Meier para a covariável Cota

3.1 Teste de Logrank

Utilizado para comparar as distribuições de sobrevivência de duas amostras, assim conseguimos comparar se existe o efeito do grupo na curva de sobrevivência.

H_0 : Funções de risco (taxa de falha) são iguais.

H_1 : Funções de risco são diferentes.

Tabela 3: P-valores do Teste de Logrank

Variavel	Pvalores
ANO_INGRESSO	<0.001
ID_SEXO	<0.001
ID_ETNIA	<0.001
INGRESSO	<0.001
COTA	<0.001
CHAMADA	<0.001
NOME_CURSO_AJUSTADO	<0.001
DURACAO	0.03

Utilizando o teste de Logrank com nível de significância de 5%, na Tabela 3 temos que todos os valores p-valores são menores que 0.05, assim há evidência para rejeitar a hipótese nula e podemos dizer que há diferenças entre os grupos. **No entanto o teste funciona bem sobre o pressuposto de riscos proporcionais, o que não ocorre, esse também é um pressuposto da Regressão de Cox, assim não utilizaremos esse métodos, no entanto, possui estimação de parâmetros próximas ao do Modelo de Tempo de Vida Acelerado.**

4 Seleção do modelo

Nessa seção buscamos o melhor modelo paramétrico, para que possibilidade adicionar covariáveis para a modelagem da sobrevivência, dado que nas seções anteriores temos inúmeras sugestões de que as covariáveis interferem no tempo de sobrevivência. Com as figuras e teste de hipótese que serão apresentadas verificaremos qual o modelo paramétrico mais adequado ao conjunto de dados.

Primeiro vamos verificar o gráfico de linearização dos modelos paramétricos, buscando o modelo que possui os pontos, mais próximos a uma reta imaginária, que se assemelharia com a bissetriz ($y = x$). Isto pois estamos verificando se a linearização do modelo paramétrico, condiz com uma linearização da sobrevivência estimada por Kaplan-Meier, assim o modelo com a linearização da sobrevivência estimada mais adequada, resultará em um melhor ajuste do modelo paramétrico. Na Figura 11, apenas é possível notar que o Modelo Exponencial é o que possui os pontos em menor linearização.

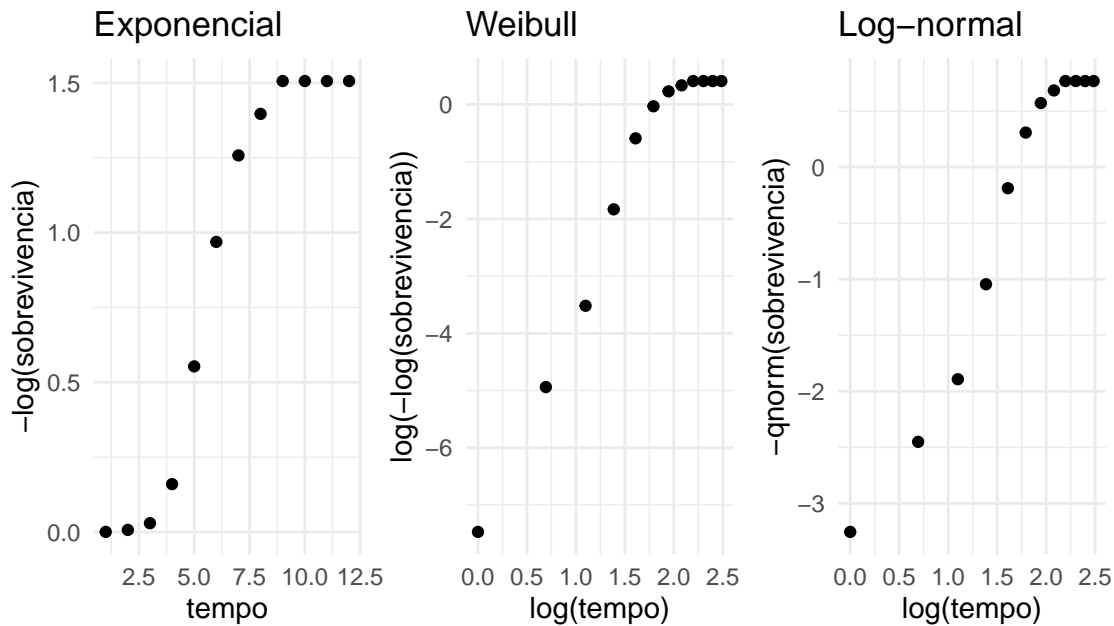


Figura 11: Linearização do modelos paramétricos para o dados da Formação dos alunos do CT

O método gráfico da sobrevivência estimada por Kaplan-Meier versus sobrevivência dos modelos paramétricas, também busca o modelo paramétrico mais adequado, pensando novamente na linearização. Analisando se a sobrevivência dos modelos probabilísticos se assemelha a do Kaplan-Meier pela linearização, pois se a relação com eles fosse dada pela bissetriz ($y = x$), teríamos que eles são iguais, então buscamos a linearidade que é dada pela bissetriz (que é a reta a 45°). Na Figura 12, novamente só podemos notar que o Modelo Exponencial é o que possui os pontos menos lineares, formando uma sigmóide.

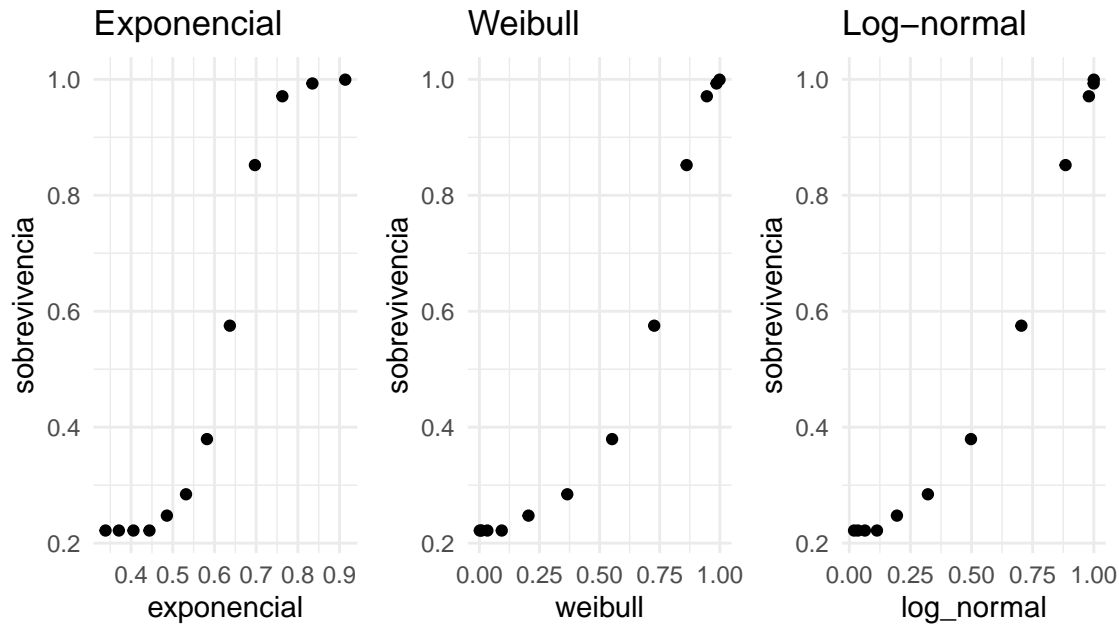


Figura 12: Sobrevivência de Kaplan-Meier vs Sobrevivência dos modelos paramétricos da Formação dos alunos do CT

No entanto também, podemos utilizar um teste de hipótese para reduzir a subjetividade da análise gráfica. Assim usaremos o teste de razão de verossimilhanças em modelos encaixados e utilizaremos a Gama Generalizada (Ou Weibull) como o modelo mais geral.

H_0 : O modelo testado é mais adequado que o modelo Generalizado.

H_1 : O modelo testado não é mais adequado que o modelo Generalizado.

Tabela 4: Seleção do modelo

Comparacoes	pvalores
Exponencial - Gamma Generalizada	0.00
Weibull - Gamma Generalizada	0.00
Log-Normal - Gamma Generalizada	0.06
Exponencial - Weibull	0.00

Utilizando o teste de razão de verossimilhança com nível de significância de 5%, na Tabela 4, temos que o único modelo que não há evidência contra H_0 é para o modelo Log-Normal, assim o escolheremos para a modelagem.

4.1 Seleção de covariáveis

Para seleção de covariáveis no modelo foi considerado os passos sugeridos por [Collett(2003)], começando com a análise individual das covariáveis, também utilizando o teste de razão de verossimilhanças em modelos encaixados e assim a Tabela 5 confirma o que o teste de Logrank e a análise descritiva sugeriram, todas as covariáveis aumentam a explicação do modelo.

Tabela 5: Testes de significância individual para cada variável

Variaveis	Pvalores
ANO_INGRESSO	<0.001
ID_SEXO	<0.001
ID_ETNIA	<0.001
INGRESSO	<0.001
COTA	<0.001
CHAMADA	<0.001
NOME_CURSO_AJUSTADO	<0.001
DURACAO	<0.001

Agora iremos ajustar um modelo com todas as covariáveis, descrito pela Tabela 8 (apêndice), no entanto é possível notar que muitas covariáveis são não significativas a 5%, assim seguiremos os próximos passos de Collett (2003) e encontraremos o modelos com todas as variáveis significativas, o que será feito agrupando covariáveis, mudando a classe base e retirando variáveis que explicam variabilidades semelhantes.

Começamos o ajuste do modelo mudando a classe base das variáveis categóricas, assim tornamos as categorias “Sem informação”, a base de todas em que ela existe, no entanto, para a variável COTA, a categoria “Universal” se mostrou uma escolha melhor à “Sem informação”, para as variáveis INGRESSO e NOME_CURSO_AJUSTADO, as classes bases são SiSU e Engenharia de Telecomunicações, respectivamente.

Ademais, foi necessário o agrupamento de algumas classes, algumas com justificativa e interpretação clara, como podemos perceber pela análise descritiva e das curvas de sobrevivência, outras as suas características de igualdade poderiam ser estudadas com mais cuidado, como a categoria Engenharia Química, Sanitária, Acústica e Computação, que a categoria que engloba os cursos de Engenharia Química, Sanitária e Ambiental, Acústica e Engenharia da Computação, assim as novas categorias são:

- **NOME_CURSO_AJUSTADO**

- Engenharia de Produção ou Controle e Automação = Engenharia de Produção + Engenharia de Controle e Automação;
- Engenharia Química, Sanitária, Acústica e Computação = Engenharia Química + Engenharia Sanitária e Ambiental + Engenharia Acústica + Engenharia da Computação
- Ciência da Computação ou Sistemas de Informação = Sistemas de Informação/CT + Ciência da Computação

- **ID_ETNIA**

- Outros = Sem informação + Indígena + Amarela + Parda

- **INGRESSO**

- Outros = Reingresso + Seleção + SiSU + Mobilidade Acadêmica
- Refugiados ou Transferência = Refugiados + Transferência
- Vestibular, Convênios e PS = Vestibular + Convênios + Processo Seletivo Seriado

- **COTA**

- PCD = Social e PCD + Racial e PCD + Social, Racial e PCD + PCD
- Social e Racial = Social e Racial + Racial + Social
- Universal = Sem informação + Universal + Escola Pública

Assim nota-se que temos pequenos ajustes, quanto aos cursos, para as etnias precisou-se agrupar as etnias com menor concentração, com exceção da Etnia Preta, para a forma de ingresso e as cotas, tiveram que ser reduzidas para 3 categorias, as cotas ficaram de formas fáceis de serem interpretadas, cotas PCD, cotas Sociais e/ou Raciais e cotas com menos restrição. Para essas novas categorias, temos que as categorias bases são Universal (para COTA), Outros (para ID_ETNIA), Outros (para INGRESSO) e continua Engenharia de Telecomunicações para

NOME_CURSO_AJUSTADO, assim como podemos ver na Tabela 6, todas as covariáveis são significativas a 5%, assim conseguimos um modelo candidato podemos ir para a análise de resíduos.

Tabela 6: Coeficientes e PValores das covariáveis ajustadas no modelo

	Betas	Pvalores
(Intercept)	2.119	< 0.001
ID_SEXOFeminino	-0.097	< 0.001
COTAPCD	0.154	0.04380
COTASocial e Racial	0.083	< 0.001
INGRESSOREfugiados ou Transferência	-0.333	< 0.001
INGRESSOVestibular, Convênios e PS	-0.133	< 0.001
ID_ETNIABranca	-0.096	< 0.001
ID_ETNIAPreta	0.132	0.00244
CHAMADACHamada	0.155	< 0.001
CHAMADAListão	0.079	< 0.001
NOME_CURSO_AJUSTADOArquitetura e Urbanismo/CT	-0.136	0.03944
NOME_CURSO_AJUSTADOCiência da Computação ou Sistemas de Informação	-0.337	< 0.001
NOME_CURSO_AJUSTADOEngenharia Química, Sanitária, Acústica e Computação	-0.158	0.01078
NOME_CURSO_AJUSTADOEngenharia Aeroespacial	-0.202	0.00689
NOME_CURSO_AJUSTADOEngenharia Civil	-0.262	< 0.001
NOME_CURSO_AJUSTADOEngenharia de Produção ou Controle e Automatização	-0.153	0.01512
NOME_CURSO_AJUSTADOEngenharia Elétrica/CT	-0.241	< 0.001
NOME_CURSO_AJUSTADOEngenharia Mecânica/CT	-0.244	< 0.001
Log(scale)	-1.128	< 0.001

Nota-se que a principal diferença de Tabela 6 para Tabela 8, é a exclusão das covariáveis DURACAO e ANO_INGRESSO, pois tem correlação alta com NOME_CURSO_AJUSTADO, COTA e INGRESSO, respectivamente. No entanto, essas covariáveis trazem mais informações, assim ficamos apenas com covariáveis categóricas.

5 Análise de Resíduo

Para verificar a qualidade do modelo proposto, precisamos avaliar a sua adequação, onde usualmente é adequado a análise gráfica, assim como na análise de adequação de modelos lineares normais e lineares generalizadas, assim utilizaremos os resíduos e_i^* , Cox-Snell, Padronizado e Deviance.

Começaremos nossa análise de resíduo com uma variação do resíduo padronizado o resíduo e_i^* , parecido com as técnicas gráficas de adequação dos modelos paramétricas, temos o gráfico da sobrevivência do resíduo pela Log-normal

e Kaplan-Meier, estão adequados pois estão muito próximos a uma reta linear, no gráfico a direita da Figura 13, a curva de sobrevivência da Log-normal tá estimando próximo ao Kaplan-Meier a sobrevivência do resíduo.

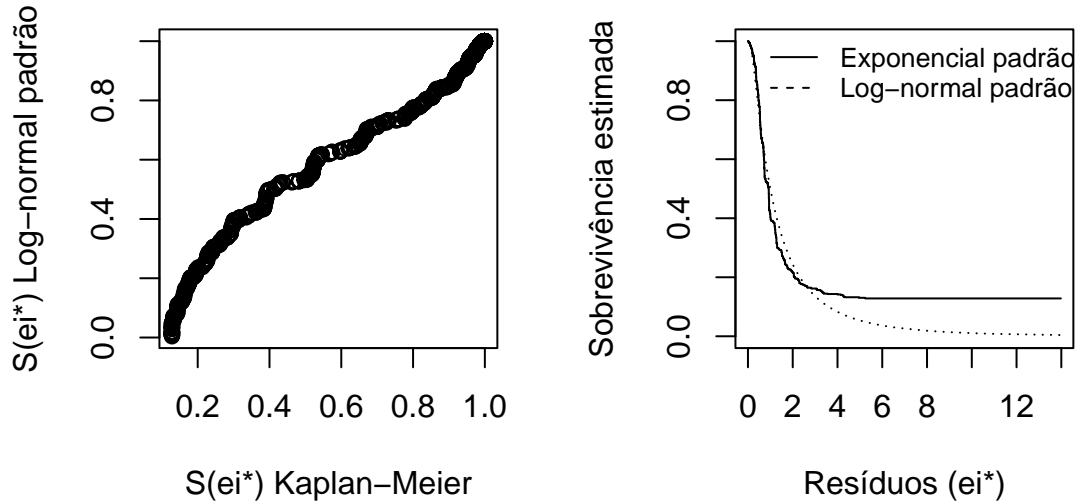


Figura 13: Sobrevivência do resíduo ei^* por Kaplan-Meier e Log-normal (à esquerda) e a Sobrevivência a estimada versus o resíduo ei^* (à direita)

Assim como a Figura 13, temos que os gráficos de Cox-Snell na Figura 14, são muito semelhantes, indicando adequação do modelo, a principal diferença para o resíduo e_{i*} é que realizamos a comparação com a sobrevivência de Kaplan-Meier, agora realizamos com a sobrevivência da exponencial padrão, pois os resíduos de Cox-Snell quando estão adequados, seguem distribuição exponencial e isso que conseguimos verificar pela análise gráfica, o modelo de Log-normal estima de forma aproximada a distribuição Exponencial padrão do resíduo de Cox-Snell.

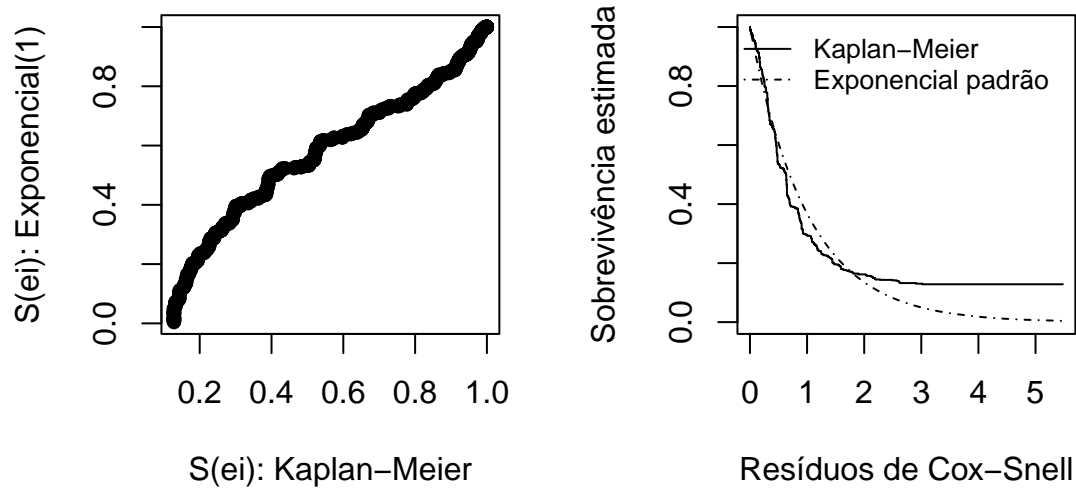


Figura 14: Sobrevivência do resíduo de Cox-Snell por Kaplan-Meier e Exponencial (1) (à esquerda) e a Sobrevivência a estimada pela Exponencial versus o resíduo de Cox-Snell (à direita)

Assim como a Figura 13, temos que os gráficos de Cox-Snell na Figura 14, são muito semelhantes, indicando adequação do modelo, a principal diferença para o resíduo e_{i*} é que realizamos a comparação com a sobrevivência de Kaplan-Meier, agora realizamos com a sobrevivência da exponencial padrão, pois os resíduos de Cox-Snell quando estão adequados, seguem distribuição exponencial e isso que conseguimos verificar pela análise gráfica, o modelo de Log-normal estima de forma aproximada a distribuição Exponencial padrão do resíduo de Cox-Snell, que como é utilizado para avaliação da adequação geral do modelo, temos bons indícios de adequação.

Finalizamos a análise de resíduo com a Figura 15, que contém os gráficos do resíduo Deviance e Padronizado, temos que os resíduos estão em torno de 0, sem valores que achatem o gráfico, indicando algum ponto de influência ou má adequação do modelo, no entanto, o comportamento não é totalmente aleatório, pois temos algumas curvas de pontos que se repetem constantemente nos dois gráficos. Para a avaliação de pontos influentes foi realizado a análise de DFBETAS para todas as covariáveis e em todos não foi encontrando nenhum desvios da regularidade. Assim iremos para a interpretação dos coeficientes ajustados.

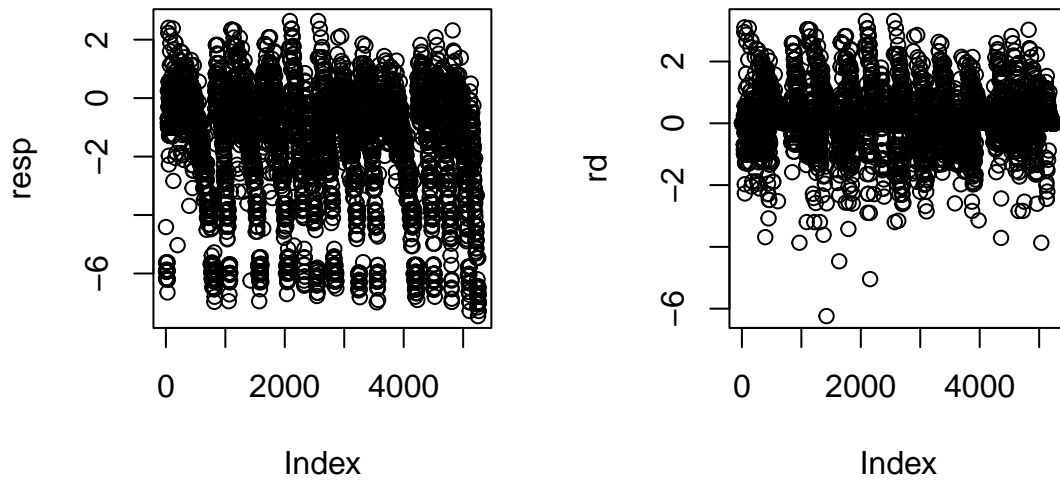


Figura 15: Resíduo Padronizado e Deviance

6 Interpretação

A parte mais importante de um modelo focado em inferência, é sua interpretação, assim necessitamos corrigir umas das principais limitações dos modelos de tempo de vida acelerado, a sua interpretação, que precisa ser feito em relação a uma taxa do tempo mediano de sobrevida.

Tabela 7: Coeficientes e PValores das covariáveis ajustadas no modelo

	Value	taxa
(Intercept)	2.119	8.320
ID_SEXOFeminino	-0.097	0.908
COTAPCD	0.154	1.166
COTASocial e Racial	0.083	1.086
INGRESSORrefugiados ou Transferência	-0.333	0.717
INGRESSOVestibular, Convênios e PS	-0.133	0.875
ID_ETNIABranca	-0.096	0.908
ID_ETNIAPreta	0.132	1.141
CHAMADACHamada	0.155	1.167
CHAMADAListão	0.079	1.082
NOME_CURSO_AJUSTADOArquitetura e Urbanismo/CT	-0.136	0.873
NOME_CURSO_AJUSTADOCiência da Computação ou Sistemas de Informação	-0.337	0.714
NOME_CURSO_AJUSTADOEngenharia Química, Sanitária, Acústica e Computação	-0.158	0.854
NOME_CURSO_AJUSTADOEngenharia Aeroespacial	-0.202	0.817
NOME_CURSO_AJUSTADOEngenharia Civil	-0.262	0.770
NOME_CURSO_AJUSTADOEngenharia de Produção ou Controle e Automatização	-0.153	0.858
NOME_CURSO_AJUSTADOEngenharia Elétrica/CT	-0.241	0.786
NOME_CURSO_AJUSTADOEngenharia Mecânica/CT	-0.244	0.783
Log(scale)	-1.128	0.324

Com as taxas dadas Tabela 7, realizaremos a interpretação para cada covariável/categoria:

- O tempo mediano para as estudantes do sexo feminino se formarem é 10% menor que estudantes do sexo masculino;
- O tempo mediano para as estudantes com cota PCD ou Social e Racial, é 17% e 9%, maior que estudantes com cota de Escola Pública, Universal ou Sem Informação;
- O tempo mediano para brancos se formarem é 10% menor que Índigenas, Amarelos e Sem informação, enquanto para pretos é 14% maior;
- Para discentes que ingressaram como Refugiados ou por Transferência o tempo para a conclusão da graduação é 30% menor e para ingressantes por Vestibular, Convênios e Processo Seletivo Seriado é 13% menor que discentes que ingressaram pelo SiSU, Reingresso ou outras formas de ingressar;
- Alunos de todos os cursos do CT tem de 30% a 13% menos tempo para se formar que alunos do curso de Engenharia de Telecomunicações, com Ciência da Computação ou Sistemas de Informação sendo os cursos com menor tempo em comparação e Arquitetura e Urbanismo o maior tempo.

7 Conclusão

A análise do tempo para formação dos alunos do CT, reitera ideias intuitivas e estereótipos, mas também traz informações novas. Mesmo com a maioria exarcebada de homens no Centro de Tecnologia, as mulheres levam menos tempo para se formar, o que pode ser objeto de estudo de alguma característica do centro para com as mulheres ou da sociedade, pois mesmo com um ambiente teoricamente pouco acolhedor, no indicador de desempenho sendo o tempo para formação, as mulheres possuem resultados superiores.

No entanto, como já foi mencionado, alguns estereótipos foram confirmados, como por exemplo alunos brancos tendo maior facilidade para se formar e alunos pretos com maior dificuldade, o que também pode ser resultado de ser uma minoria ainda menor que as das mulheres. Para os cotistas, temos que os alunos com PCD ou com cota Social e/ou Racial, possuem maior tempo para formação que alunos com cota de Escola Pública ou Universal. Para a análise da forma de ingresso, o estudo se torna um pouco mais complicado, pelas enormes associações de fatores, no entanto SiSu e Vestibular formam mais de 80% das observações e SiSU fica no grupo com maior tempo para se formar, o que em boa parte deve-se a uma quantidade alta de dados censurados.

Nesse ínterim, destaca-se da variável INGRESSO que discentes que ingressaram como transferência ou refugiados possuem o menor tempo para formação. Assim, acredito que esse estudo trouxe muitas confirmações do imaginário público, mas também trouxe conhecimentos novos, que se observadas com cuidado ajudam a focar nos problemas concretos. Para uma próxima análise, deve-se considerar o ajuste de modelos com riscos competitivos.

Apêndice - Código R

```
library(dplyr)
library(ggplot2)
library(survival)
library(flexsurv)
library(magrittr)
library(ggfortify)
library(patchwork)

grafico <- function(data, v1, title, v2 = tempo){
  ggplot(data) +
    geom_point(aes({v2},{v1})) +
    labs(title = title)
}

bar_plot <- function(df, v) {
  mycolors = c(RColorBrewer::brewer.pal(name="Set2", n = 8),
    RColorBrewer::brewer.pal(name="Paired", n = 6))
  ggplot(df, aes(
    x = {{ v }},
    y = prop.table(stat(count)),
    fill = {{ v }},
    label = scales::percent(prop.table(stat(count)))
  )) +
    geom_bar(position = "dodge") +
    geom_text(
      stat = "count",
      position = position_dodge(.9),
      vjust = -0.3,
      size = 3.5
    ) +
```

```

    scale_y_continuous(labels = scales::percent) +
    scale_fill_manual(values = mycolors) +
    labs(y = "Proporção", x = df %>%
      select({{ v }}) %>%
      names()) +
    theme(plot.title = element_text(hjust = 0.5, size = 10),
          legend.position="none")
  }

bar_plot2 <- function(df, v) {
  mycolors = c(RColorBrewer::brewer.pal(name="Set2", n = 8),
    RColorBrewer::brewer.pal(name="Paired", n = 6))
  ggplot(df, aes(
    x = {{ v }},
    y = prop.table(stat(count)),
    fill = {{ v }},
    label = scales::percent(prop.table(stat(count)))
  )) +
    geom_bar(position = "dodge") +
    scale_y_continuous(labels = scales::percent) +
    scale_fill_manual(values = mycolors) +
    labs(y = "Proporção", x = df %>%
      select({{ v }}) %>%
      names()) +
    theme(plot.title = element_text(hjust = 0.5, size = 10),
          legend.position="none")
  }

significancia_covariaveis <- function(df, dist, ...){
  mc <- match.call(expand.dots = FALSE)
  call_args <- mc$...
  name <- call_args |> unlist()
  fit<-survreg(Surv(TEMPO,CENSURA)~1, data = df, dist=dist)
  pvalores = matrix(ncol = 2, nrow = length(name)) |> data.frame() |>
    select(Variaveis = X1, Pvalores = X2)
  j = 1
  for(i in name){
    v1 = df |> select( {{i}} ) |> unlist()
    fit_covariavel<-survreg(Surv(TEMPO,CENSURA)~v1, data = df, dist= dist)
    pvalores[j,1] = deparse(i)
    pvalores[j,2] = 1-pchisq(2*(fit_covariavel$loglik[2]-fit$loglik[2]),1)
    j = j +1
  }
  pvalores
}

#survdif(Surv(TEMPO,CENSURA)~ID_SEXO, data = df,rho=0) -> a
df = readr::read_csv('sobrevivencia.csv')
df |> select(-ANO_EVASAO, - SIGLA_CENTRO, - TURNO_CURSO, -SITUACAO) |>

```

```

rename(SEXO = ID_SEXO, ETNIA = ID_ETNIA, CURSO = NOME_CURSO_AJUSTADO,
        INICIO = ANO_INGRESSO) |>
slice(2:6) |>
mypdf1::pdf1_tbl("Observações Iniciais do Banco") |>
kableExtra::add_footnote(c("Algumas variáveis sofreram alterações no nome para melhorar visualização da ta
mypdf1::pdf1_summary(df) |>
mypdf1::pdf1_tbl("Descrição das Variáveis Não-categóricas") |>
kableExtra::add_footnote(c("Variáveis inicialmente tratadas como numérica pelo SoftWare R"))
df = df |> mutate(INGRESSO = as.factor(INGRESSO), COTA = as.factor(COTA),
                  CHAMADA = as.factor(CHAMADA),
                  NOME_CURSO_AJUSTADO = as.factor(NOME_CURSO_AJUSTADO),
                  TURNO_CURSO = as.factor(TURNO_CURSO),
                  ID_SEXO = forcats::fct_recode(factor(ID_SEXO),
                                                'Masculino' = '1',
                                                'Feminino' = '2'),
                  ID_ETNIA = forcats::fct_recode( factor(ID_ETNIA),
                                                'Branca' = '1', 'Preta' = '2',
                                                'Amarela' = '3', 'Parda' = '4',
                                                'Indígena' = '5',
                                                'Sem informação' = '6'))
(df |> mutate(ANO_INGRESSO = as.factor(ANO_INGRESSO))) |>
bar_plot(ANO_INGRESSO) + theme(legend.position="none")) /
(bar_plot(df) |>
  filter(ANO_EVASAO == ANO_EVASAO) |>
  mutate(ANO_EVASAO = as.factor(ANO_EVASAO)), ANO_EVASAO) +
  theme(legend.position="none"))
bar_plot(df, ID_SEXO) + bar_plot(df |>
  mutate(ID_ETNIA =
    forcats::fct_infreq(ID_ETNIA)), ID_ETNIA) +
  theme(axis.text.x = element_text(angle = 20, size = 10, hjust = 1))
bar_plot2(df |> mutate(INGRESSO = forcats::fct_infreq(INGRESSO)), INGRESSO) +
  coord_flip() +
  bar_plot2(df |> mutate(COTA = forcats::fct_infreq(COTA)), COTA) + coord_flip()
bar_plot(df, CHAMADA) +
  bar_plot(df |> mutate(DURACAO = as.factor(DURACAO)), DURACAO)
bar_plot2(df |>
  mutate(NOME_CURSO_AJUSTADO =
    forcats::fct_infreq(NOME_CURSO_AJUSTADO)),
  NOME_CURSO_AJUSTADO) + coord_flip()
df = df |> filter(TEMPO != 0, SITUACAO != 'Desistência')
fit_ekm <- survfit(Surv(TEMPO, CENSURA)~1, data = df)
autoplot(fit_ekm, ylab='Porcentagem de conclusão', xlab='Tempo')
fit_ekm <- survfit(Surv(TEMPO, CENSURA)~ID_SEXO, data = df)
p1 <- autoplot(fit_ekm, ylab='Porcentagem de conclusão') +
  theme(legend.title = element_blank())
fit_ekm <- survfit(Surv(TEMPO, CENSURA)~CHAMADA, data = df)
p2 <- autoplot(fit_ekm, ylab='Porcentagem de conclusão', xlab='Tempo')+
  theme(legend.title = element_blank())
p1 + p2
fit_ekm <- survfit(Surv(TEMPO, CENSURA)~ID_ETNIA, data = df |>
  filter(!ID_ETNIA %in% c('Amarela', 'Indígena',

```

```

                                'Sem informação'))))
p1 = autoplot(fit_ekm, ylab='Porcentagem de conclusão', xlab='Tempo') +
  theme(legend.title = element_blank())
fit_ekm <- survfit(Surv(TEMPO,CENSURA)~ID_ETNIA, data = df |>
  filter(ID_ETNIA %in% c('Amarela', 'Indígena',
                        'Sem informação'))))
p2 = autoplot(fit_ekm, ylab='Porcentagem de conclusão', xlab='Tempo') +
  theme(legend.title = element_blank())
p1 + p2
fit_ekm <- survfit(Surv(TEMPO,CENSURA)~INGRESSO, data = df |>
  filter(INGRESSO %in% c('SiSU', 'Vestibular', 'Reingresso',
                        'Processo Seletivo Seriado')) |>
  mutate(INGRESSO = forcats::fct_recode(INGRESSO, 'PEIES' =
                                         "Processo Seletivo Seriado"))))
p1 = autoplot(fit_ekm, ylab='Porcentagem de conclusão', xlab='Tempo') +
  theme(legend.title = element_blank())
fit_ekm <- survfit(Surv(TEMPO,CENSURA)~INGRESSO, data = df |>
  filter(INGRESSO %in% c('Transferências', 'Convênios',
                        'Refugiados'))))
p2 = autoplot(fit_ekm, ylab='Porcentagem de conclusão', xlab='Tempo') +
  theme(legend.title = element_blank())
p1 + p2
fit_ekm <- survfit(Surv(TEMPO,CENSURA)~COTA, data = df |>
  filter(COTA %in% c('Universal', 'Escola Pública',
                    'Racial', 'Social'))))
p1 = autoplot(fit_ekm, ylab='Porcentagem de conclusão', xlab='Tempo') +
  theme(legend.title = element_blank())
fit_ekm <- survfit(Surv(TEMPO,CENSURA)~COTA, data = df |>
  filter(COTA %in% c('PCD',
                    'Sem informação',
                    'Social e Racial'))))
p2 = autoplot(fit_ekm, ylab='Porcentagem de conclusão', xlab='Tempo') +
  theme(legend.title = element_blank())
p1 + p2
data.frame(Variavel = c('ANO_INGRESSO', 'ID_SEXO', 'ID_ETNIA', 'INGRESSO', 'COTA',
                        'CHAMADA', 'NOME_CURSO_AJUSTADO', 'DURACAO'),
           Pvalores = c(2e-16, 3e-10, 2e-16, 2e-16, 2e-16, 2e-16, 0.03)) |>
  mutate(Pvalores = format.pval(Pvalores, 3, eps = 0.001)) |>
  mypdf1::pdf1_tbl("P-valores do Teste de Logrank")
fitg<-flexsurvreg(Surv(TEMPO,CENSURA)~1, data = df, dist='gengamma')
fite<-survreg(Surv(TEMPO,CENSURA)~1, data = df, dist='exponential')
fitw<-survreg(Surv(TEMPO,CENSURA)~1, data = df, dist='weibull')
fitln<-survreg(Surv(TEMPO,CENSURA)~1, data = df, dist='lognormal')
p.valore= 1-pchisq(2*(fitg$loglik[1]-fite$loglik[1]),2)
p.valorw= 1-pchisq(2*(fitg$loglik[1]-fitw$loglik[1]),1)
p.valorln= 1-pchisq(2*(fitg$loglik[1]-fitln$loglik[1]),1)
p.valorew = 1-pchisq(2*(fitw$loglik[1]-fite$loglik[1]),1)
fit_inicial <- survfit(Surv(TEMPO,CENSURA)~1, data = df)
ekm_inicial <- fit_inicial$surv %>%
  data.frame(sobrevivencia=.) %>%

```

```

mutate(tempo = fit_inicial$time)
grafico(ekm_inicial, -log(sobrevivencia), "Exponencial") +
grafico(ekm_inicial, log(-log(sobrevivencia)),
        "Weibull", log(tempo)) +
grafico(ekm_inicial, -qnorm(sobrevivencia),
        "Log-normal", log(tempo))
alphae = exp(fite$coef)
alphaw = exp(fitw$coef)
gamaw = 1/fitw$scale
mu = fitln$coeff
sigma = fitln$scale
ekm_inicial <- ekm_inicial %>%
  mutate(exponencial = exp(-tempo/alphae),
         weibull = exp(-(tempo/alphaw)^gamaw),
         log_normal = pnorm((-log(tempo)+ mu)/ sigma))
grafico(ekm_inicial, sobrevivencia, "Exponencial", exponencial) +
grafico(ekm_inicial, sobrevivencia,
        "Weibull", weibull) +
grafico(ekm_inicial, sobrevivencia,
        "Log-normal", log_normal)
data.frame(Comparacoes = c('Exponencial - Gamma Generalizada',
                           'Weibull - Gamma Generalizada',
                           'Log-Normal - Gamma Generalizada',
                           'Exponencial - Weibull'),
          pvalores = c(p.valore, p.valorw, p.valorln, p.valorew)) %>%
  mypdf1::pdf1_tbl('Seleção do modelo')
significancia_covariaveis(df, 'lognormal', ANO_INGRESSO, ID_SEXO, ID_ETNIA,
                          INGRESSO, COTA, CHAMADA, NOME_CURSO_AJUSTADO,
                          DURACAO) |> mutate(Pvalores =
                                             format.pval(Pvalores, 3,
                                                         eps = 0.001)) |>
  mypdf1::pdf1_tbl("Testes de significância individual para cada variável")
df1 = df |> mutate(ID_ETNIA = factor(ID_ETNIA, c('Sem informação','Branca', 'Preta', 'Amarela',
        'Parda', 'Indígena')),
                  INGRESSO = factor(INGRESSO, c("SiSU","Refugiados","Mobilidade Acadêmica",
        "Convênios",
        "Processo Seletivo Seriado",
        "Reingresso",
        "Seleção",
        "Transferências",
        "Vestibular")),
                  COTA = factor(COTA, c("Universal", "Sem informação", "Escola Pública",
        "PCD", "Racial", "Racial e PCD",
        "Social", "Social e PCD",
        "Social e Racial",
        "Social, Racial e PCD")),
                  CHAMADA = factor(CHAMADA, c("Sem informação", "Chamada",
        "Listão")),
                  NOME_CURSO_AJUSTADO = factor(NOME_CURSO_AJUSTADO,
        c("Engenharia de Telecomunicações",
        "Arquitetura e Urbanismo/CT",

```

```

        "Ciência da Computação",
        "Engenharia Acústica",
        "Engenharia Aeroespacial",
        "Engenharia Civil",
        "Engenharia de Computação",
        "Engenharia de Controle e Automação",
        "Engenharia de Produção",
        "Engenharia Elétrica/CT",
        "Engenharia Mecânica/CT",
        "Engenharia Química",
        "Engenharia Sanitária e Ambiental/CT",
        "Sistemas de Informação/CT")) |>
mutate(ID_ETNIA = forcats::fct_recode(ID_ETNIA, "Outros" = "Sem informação",
        "Outros" = "Amarela",
        "Outros" = "Indígena",
        "Outros" = "Parda"),
INGRESSO = forcats::fct_recode(INGRESSO, "Vestibular, Convênios e PS" =
        "Convênios",
        "Outros" = "Reingresso",
        "Outros" = "Seleção",
        "Outros" = "SiSU",
        "Outros" = "Mobilidade Acadêmica",
        "Refugiados ou Transferência" =
        "Refugiados",
        "Refugiados ou Transferência" =
        "Transferências",
        "Vestibular, Convênios e PS" =
        "Vestibular",
        "Vestibular, Convênios e PS" =
        "Processo Seletivo Seriado"),
NOME_CURSO_AJUSTADO = forcats::fct_recode(NOME_CURSO_AJUSTADO,
        "Engenharia de Produção ou Controle e Automação" =
        "Engenharia de Controle e Automação",
        "Engenharia de Produção ou Controle e Automação" =
        "Engenharia de Produção",
        "Engenharia Química, Sanitária, Acústica e Computação" =
        "Engenharia Química",
        "Engenharia Química, Sanitária, Acústica e Computação" =
        "Engenharia de Computação",
        "Engenharia Química, Sanitária, Acústica e Computação" =
        "Engenharia Química, Sanitária, Acústica e Computação" =
        "Engenharia Sanitária e Ambiental/CT",
        "Engenharia Química, Sanitária, Acústica e Computação" =
        "Engenharia Acústica",
        "Ciência da Computação ou Sistemas de Informação" =
        "Ciência da Computação",
        "Ciência da Computação ou Sistemas de Informação" =
        "Sistemas de Informação/CT"
        ),
COTA = forcats::fct_recode(COTA, "PCD" = "Racial e PCD",
        "PCD" = "Social e PCD",
        "PCD" = "Social, Racial e PCD",

```



```

"Universal" = "Escola Pública",
"Social e Racial" = "Racial",
"Social e Racial" = "Social",
"Universal" = "Sem informação"))

fit1<-survreg(Surv(TEMPO,CENSURA)~ ID_SEXO + COTA + INGRESSO + ID_ETNIA +
              CHAMADA + NOME_CURSO_AJUSTADO,
              data = df1, dist='lognormal')
ajuste = summary(fit1)
ajuste$table |> as.data.frame() |>
  select(-`Std. Error`, -z) |> rename(Pvalores=p, Betas = Value) |>
  mutate(Pvalores = format.pval(Pvalores, 3, eps = 0.001)) |>
  mypdf1::pdf1_tbl("Coeficientes e PValores das covariáveis ajustadas no modelo")
X = model.matrix(~ID_SEXO + COTA + INGRESSO + ID_ETNIA + CHAMADA +
                 NOME_CURSO_AJUSTADO, data = df1)
beta = fit1$coefficients
sigmahat = fit1$scale
res = (log(df$TEMPO)-X%*(beta))/sigmahat #Modelo Log-normal
resid = exp(res)
ekm <- survfit(Surv(resid, CENSURA)~1, data = df, type=c("kaplan-meier"))
resid = ekm$time
sln = pnorm(-log(resid))
par(mfrow=c(1,2))
plot(ekm$surv,sln, lty=c(1,1), xlab="S(ei*) Kaplan-Meier",
     ylab="S(ei*) Log-normal padrão")
plot(ekm, conf.int = F, mark.time = F, xlab = "Resíduos (ei*)",
     ylab = 'Sobrevivência estimada', pch = 16)
lines(resid, sln, lty=3)
legend(0, 1.05, lty=c(1,2), legend=c("Exponencial padrão","Log-normal padrão"),
      bty="n", cex=0.8)
#Resíduos Cox-Snell
par(mfrow=c(1,2))
ei <- -log(1-pnorm(res))
ekm1 <- survfit(Surv(ei,df$CENSURA)~1)
t <- ekm1$time
st <- ekm1$surv
sexp <- exp(-t)
plot(st, sexp, xlab="S(ei): Kaplan-Meier", ylab= "S(ei): Exponencial(1)", pch=16)
plot(ekm1, conf.int = F, mark.time = F, xlab = 'Resíduos de Cox-Snell',
     ylab = 'Sobrevivência estimada')
lines(t, sexp, lty=4)
legend(-0.5, 1.03, lty=c(1,4), legend=c("Kaplan-Meier","Exponencial padrão"),
      bty="n", cex=0.8)
par(mfrow=c(1,2))
resp = (log(df$TEMPO)-X%*(beta))/sigmahat
plot(resp)

rd=resid(fit1, type="deviance")
plot(rd)
ajuste$table |> as.data.frame() |>
  select(-`Std. Error`, -z, -p) |> mutate(taxa = exp(Value)) |>

```

```

mypdf1::pdf1_tbl("Coeficientes e PValores das covariáveis ajustadas no modelo")
fit <- survreg(Surv(TEMPO, CENSURA) ~ ID_SEXO + COTA + INGRESSO + ID_ETNIA +
              CHAMADA + NOME_CURSO_AJUSTADO + DURACAO + ANO_INGRESSO,
              data = df, dist = 'lognormal')
ajuste = summary(fit)
ajuste$table |> as.data.frame() |>
  select(-`Std. Error`, -z) |> rename(Pvalores=p, Betas = Value) |>
  mutate(Pvalores = format.pval(Pvalores, 3, eps = 0.001)) |>
  mypdf1::pdf1_tbl("Coeficientes e PValores no modelo inicial com todas as covariáveis")

```

Tabela 8: Coeficientes e PValores no modelo inicial com todas as covariáveis

	Betas	Pvalores
(Intercept)	-33.787	< 0.001
ID_SEXOFeminino	-0.058	< 0.001
COTAPCD	0.167	0.02838
COTARacial	0.185	< 0.001
COTARacial e PCD	3.987	< 0.001
COTASem informação	-0.015	0.72197
COTASocial	0.034	0.19572
COTASocial e PCD	3.669	< 0.001
COTASocial e Racial	0.144	0.00545
COTASocial, Racial e PCD	3.187	< 0.001
COTAUiversal	0.011	0.50298
INGRESSOMobilidade Acadêmica	0.000	NA
INGRESSOOutros	0.000	NA
INGRESSOProcesso Seletivo	0.114	0.40080
Seriado		
INGRESSOREfugiados	2.269	< 0.001
INGRESSOREingresso	0.096	0.53780
INGRESSOSeleção	7848.258	< 0.001
INGRESSOSiSU	0.198	0.15262
INGRESSOTransferências	-0.094	0.47791
INGRESSOVestibular	0.126	0.35688
ID_ETNIAPreta	0.132	0.00550
ID_ETNIAAmarela	0.049	0.76854
ID_ETNIAParda	0.048	0.08903
ID_ETNIAIndígena	0.226	0.32513
ID_ETNIASem informação	0.081	0.02467
CHAMADAListão	-0.082	< 0.001
CHAMADASem informação	-0.118	< 0.001
NOME_CURSO_AJUSTADOCiência da Computação	-0.309	< 0.001
NOME_CURSO_AJUSTADOEngenharia Acústica	0.235	< 0.001
NOME_CURSO_AJUSTADOEngenharia Aeroespacial	-0.073	0.15621
NOME_CURSO_AJUSTADOEngenharia Civil	-0.111	< 0.001
NOME_CURSO_AJUSTADOEngenharia de Computação	0.067	0.07312
NOME_CURSO_AJUSTADOEngenharia de Controle e Automação	0.125	< 0.001
NOME_CURSO_AJUSTADOEngenharia de Produção	-0.072	0.02970
NOME_CURSO_AJUSTADOEngenharia de Telecomunicações	0.139	0.03246
NOME_CURSO_AJUSTADOEngenharia Elétrica/CT	-0.086	0.00772
NOME_CURSO_AJUSTADOEngenharia Mecânica/CT	-0.082	0.01359
NOME_CURSO_AJUSTADOEngenharia Química	-0.131	< 0.001
NOME_CURSO_AJUSTADOEngenharia Sanitária e Ambiental/CT	0.363	0.08138
NOME_CURSO_AJUSTADOSistemas de Informação/CT	-0.108	0.00546
DURACAO	-0.044	< 0.001