

Identificação dos fatores que influenciam no tempo até a evasão/ retenção dos alunos do ICEB-UFOP utilizando técnicas paramétricas em Análise de Sobrevida

Anderson Garneth de Castro¹

Graziela Dutra Rocha Gouvêa²

1 Introdução

Sabe-se que a evasão nas universidades são problemas complexos, decorrentes de várias causas e que devem ser colocados em discussão de modo permanente, para solucioná-los ou evitar que esses eventos continuem aumentando.

A evasão e a retenção observada nas universidades brasileiras vêm preocupando não só os professores e alunos, mas também as autoridades ligadas ao ensino. A realidade da UFOP não é diferente. Segundo informações da PROGRAD/UFOP, a evasão atinge índice superior a 20% e varia conforme os seus institutos. No período de 2007/2 a 2011/2, o maior índice observado foi no Instituto de Ciências Exatas e Biológicas que atingiu 37,1%.

Este trabalho tem por objetivo estudar os fatores que influenciam no tempo até a evasão/trancamento dos alunos dos cursos oferecidos pelo ICEB-UFOP, por meio de técnicas em Análise de Sobrevida, de modo que tais resultados possam ajudar a todos envolvidos com o desenvolvimento da instituição a solucionar ou ao menos minimizar tal problema.

2 Material e Métodos

Foram utilizadas técnicas paramétricas de Análise de Sobrevida para analisar quais fatores influenciam no tempo até a evasão/trancamento do estudante. Os dados reais referem-se a 2658 alunos matriculados entre 2003 e 2013 em disciplinas ministradas no ICEB/UFOP. O tempo foi medido em horas cursadas e as covariáveis estudadas foram: ANO - Ano de admissão que vale 1 se admito de 2009 até 2013 e 0 se de 2003 até 2008; SEM - Semestre de admissão que vale 1 se semestre II e 0 se semestre I; AD - Modo admissão que vale 1 se vestibular e 0 se outros; TU - Turno que vale 1 se noturno e 0 se outros; R - Cor da pele que vale 1 se branco e 0 se outros; OR - Estado de nascimento que vale 1 se Minas Gerais e 0 se outros; COD - Curso que vale 1 se da área de Ciências Exatas e 0 se Ciências

¹ Estudante de graduação em Estatística - UFOP. e-mail: adsth91@yahoo.com.br

² Professora Adjunto II do DEEST-UFOP.

Biologicas; ID - Idade (em anos); CID - Cidade de nascimento que vale 1 se Ouro Preto e 0 se outros; SEXO – que vale 1 se sexo masculino e 0 se sexo feminino.

Foi realizado o teste da razão de verossimilhança (TRV), para verificar o ajuste dos modelos Exponencial, Log-Normal e Weibull, caso particulares da Gama Generalizada, distribuição na qual o teste TRV foi baseado. Tal teste é baseado no \ln da razão entre as duas verossimilhanças: $TRV = -2[\ln L(\theta_0) - \ln L(\theta_i)] \sim \chi_g^2$; em que, g é o número de restrições (Kalbfleisch e Prentice, 2002). A hipótese foi testada com um nível de significância a 5%.

Após a verificação gráfica e a constatação pelo TRV ao nível de 5% de confiança, de qual distribuição é adequada para o ajuste dos dados, passou a seleção de covariáveis pela proposta de Collet (1994).

Verificadas as significâncias estatísticas em cada passo da proposta, os modelos com as possíveis covariáveis e interações significativas foram determinados e procedeu-se a verificação da adequação dos modelos. A verificação foi feita por meio da Análise de Resíduos estimadas pelo método de Kaplan-Meier e pelos Resíduos de Cox-Snell. (Colosimo e Giolo, 2006). A partir das análises gráficas, o modelo final foi determinado e procederam-se as interpretações das covariáveis e termos de interações que compõem tal modelo.

Todas as rotinas que geraram os resultados para análise dos dados foram realizadas no Software Estatístico R.

3 Resultados e Discussões

Como pode ser visto na Figura 1, foram construídos gráficos das estimativas das sobrevivências pelo método de Kaplan-Meier versus as estimativas das sobrevivências obtidas a partir dos modelos exponencial, Weibull e log-normal. Com base nesta Figura, foi possível verificar que o modelo exponencial não se ajusta aos dados, pois a curva se apresenta mais afastada da reta do que os modelos Weibull e log-normal, o que sugere estas distribuições podem ser apropriadas para descrever o tempo até a evasão/trancamento.

Para confirmar tais resultados, foram construídos gráficos linearizados para os modelos (Figura 2). Os resultados reforçam a indicação dos modelos de Weibull e log-normal como possivelmente adequados para os dados sob estudo.

Também foram geradas as curvas de sobrevivência estimadas por meio do ajuste de ambos os modelos versus a curva de sobrevivência estimada por Kaplan-Meier e a partir dos resultados gráficos foi possível verificar que ambos os modelos apresentaram ajustes satisfatórios.

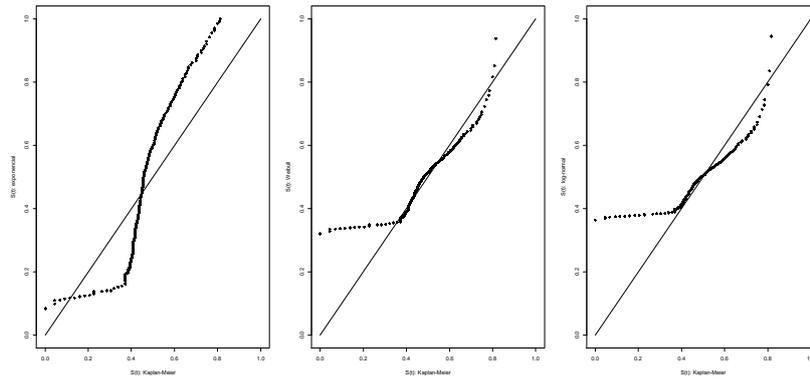


Figura 5: Gráficos das sobrevivências estimadas por Kaplan-Meier versus as sobrevivências estimadas pelos modelos exponencial, de Weibull e log-normal.

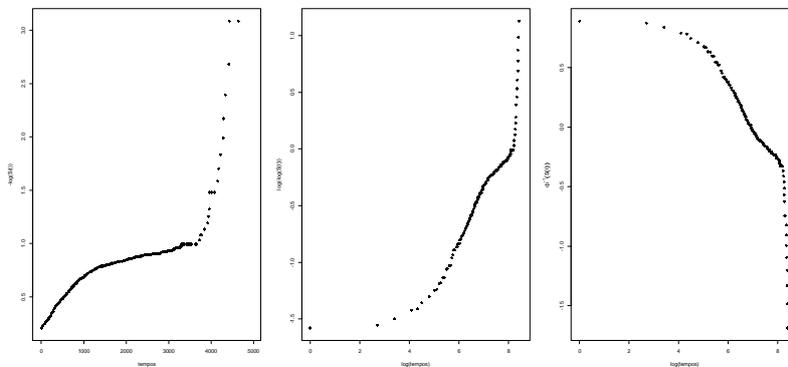


Figura 6: Gráficos linearizados do modelo exponencial, Weibull, e log-normal.

A Tabela 2 apresenta os Testes da Razão de Verossimilhanças para os modelos estudados e a partir destes resultados, foi possível verificar que o modelo que melhor se ajustou aos dados em estudo foi o modelo log-normal.

Tabela 2: Logaritmo da função de verossimilhança e resultados dos TRV.

Modelo	Log(L(θ))	TRV	P-Valor
Gama Generalizado	10666,52	-	-
Exponencial	12317,20	3301,36	<0,01
Weibull	10675,30	17,56	<0,01
Log-normal	10666,60	0,16	0,69

O modelo final, que envolve as covariáveis e interação ANO SEM TU R OR COD SEXO ANO TU*COD foi determinado. Foram utilizadas análises de resíduos para confirmar a adequação do modelo. Os gráficos das sobrevivências dos resíduos estimadas por Kaplan-

Meier e pelo modelo log-normal padrão bem como os gráficos de suas respectivas curvas de sobrevivência estimadas encontram-se apresentados na Figura 13 e 19 respectivamente.

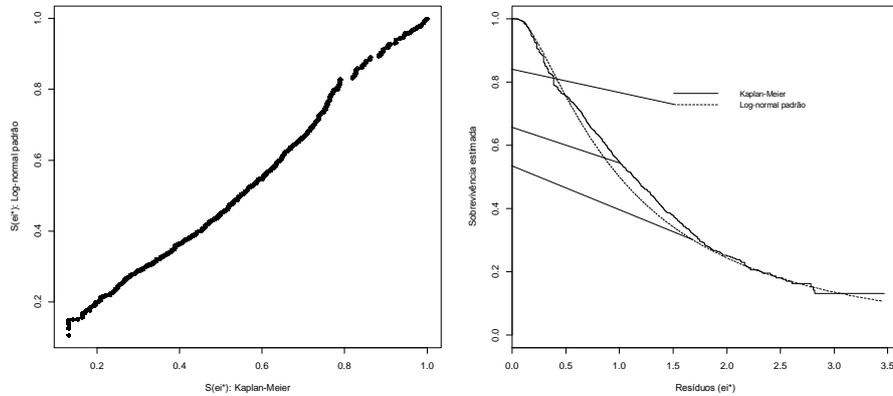


Figura 13 - Sobrevivências dos resíduos estimadas pelo método de Kaplan-Meier e pelo modelo log-normal padrão e respectivas curvas de sobrevivência estimadas.

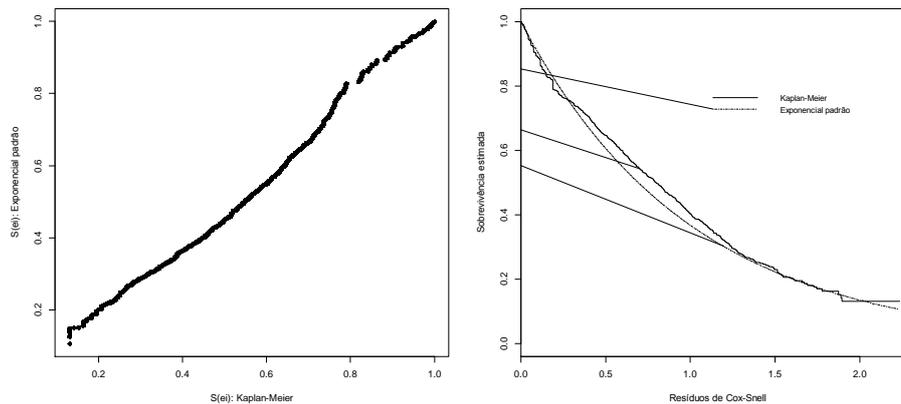


Figura 19: Sobrevivência dos resíduos Cox-Snell estimadas pelo método de Kaplan Meier e exponencial padrão (à esquerda) e respectivas curvas de sobrevivência estimadas (à direita)

Assim, o modelo final ficou determinado pelos efeitos principais identificados inicialmente e os termos de interação significativos, como pode ser visto na Tabela 3.

Tabela 3: Estimativas dos parâmetros dos modelos de regressão log-normal.

Covariável	Estimativa	E.P	p-valor
Constante	9,043	0,3885	<0,01
ANO	-2,358	0,2181	<0,01
SEM	2,082	0,2449	<0,01

TU	0,888	0,4414	0,04
R	-0,394	0,1878	0,03
OR	-1,388	0,2586	<0,01
COD	-0,523	0,3755	0,1
SEXO	-1,095	0,1947	<0,01
TU*COD	-1,269	0,4948	0,01

4 Conclusões

Conclui-se, após a análise paramétrica, que em relação ao tempo mediano até a evasão/trancamento dos discentes do ICEB/UFOP: aqueles que foram admitidos entre os anos de 2003 a 2008 possuem tempo maior (aproximadamente 10 vezes) do que daqueles admitidos entre os anos de 2009 a 2013. Os discentes que ingressaram no segundo semestre possuem tempo maior (aproximadamente 8 vezes) do que aqueles admitidos no primeiro semestre. Quanto ao turno conclui-se que o tempo para os alunos que estudam em cursos noturnos é maior (aproximadamente 2,4 vezes) do que dos outros turnos. Alunos que declararam cor da pele branca possuem tempo menor (aproximadamente a metade) do que aqueles que se declararam não brancos. Em relação ao estado de origem, os discentes oriundos de Minas Gerais possuem tempo menor do que aqueles de outros estados. O tempo mediano até a evasão/trancamento dos discentes matriculados em curso da área de exatas é aproximadamente a metade daqueles matriculados em curso da área de ciências biológicas. Quanto ao sexo, o tempo dos estudantes do sexo feminino é maior (aproximadamente 3 vezes) do que o masculino. Conclui-se ainda que o tempo dos discentes que fazem curso na área de ciências biológicas e estudam em turno que não seja o noturno é três vezes maior do que daqueles matriculados em curso da área de exatas no turno noturno.

5 Referências Bibliográficas

- [1] COLLETT, D. (1994). **Modelling Survival Data in Medical Reserch**. Chapman and Hall, London.
- [2] COLOSIMO, E. A.; GIOLO, S. R. **Análise de Sobrevivência Aplicada**. São Paulo: Edgard Blücher, 2006. 369p.
- [3] KALBFLEISCH, J. D.; PRENTICE, R. L. **The statistical analysis of failure time data**. 2. ed. New York: J. Wiley; 2002. 437 p.