



ANÁLISE DO PERFIL DOS ALUNOS EGRESSOS DO CURSO DE ESTATÍSTICA DA UNICAMP

Palavras-chave: Egressos; Bacharelado em Estatística; Regressão Logística.

Autores:

Vanessa Aparecida Fernandes, IMECC – Unicamp
Prof. Dr. Rafael Pimentel Maia, IMECC – Unicamp

INTRODUÇÃO

No Brasil, enquanto a demanda por estatísticos cresce cada vez mais, apenas um pouco mais de quatro centenas desses profissionais se formam anualmente¹ e, mesmo assim, como o acompanhamento de egressos está em fase inicial (SIMON; PACHECO, 2017), não há muitas pesquisas sobre as trajetórias ocupacionais desses estatísticos. Além disso, embora vários estudos indiquem que a formação em nível superior é um dos fatores que mais explicam as oportunidades no mercado de trabalho (RIBEIRO; SCHLEGEL, 2015; SAITO, 2006), as trajetórias ocupacionais dos egressos e os ganhos salariais deles também dependem de características socioeconômicas e demográficas (VAZ, 2022).

Nesse contexto, este trabalho tem por objeto analisar o perfil dos alunos egressos entre 2005 e 2016 do curso de Bacharelado em Estatística da Unicamp e avaliar suas possíveis trajetórias, como o ingresso em uma pós-graduação *stricto sensu* (mestrado e doutorado) e/ou o ingresso no mercado de trabalho, bem como estudar características sociodemográficas e de desempenho acadêmico que possam estar associadas a cada um dos casos. Para tanto, além de uma análise descritiva, será ajustado um modelo de regressão logística para estimar a chance de um aluno concluinte ingressar em um programa de pós-graduação e um modelo aditivo generalizado de localização, dispersão e forma para estimar o ganho salarial médio dos egressos que foram para o mercado de trabalho.

MATERIAIS

Para este estudo foram mobilizados registros de quatro bancos de dados de alunos egressos do curso de Bacharelado em Estatística nos anos de 2005 a 2016: 1) Dados de desempenho acadêmico dos estudantes no egresso, disponibilizados pela Diretoria Acadêmica da Unicamp (DAC); 2) Dados socioeconômicos e demográficos dos estudantes no momento da inscrição no vestibular, disponibilizados pela Comissão Permanente para os Vestibulares da Unicamp (Comvest); 3) Dados referentes à atividade trabalhista a partir da Relação Anual de Informações Sociais (RAIS) do Ministério da Economia; e 4) Dados referentes ao ingresso na pós-graduação *stricto sensu* da base pública da CAPES.

As bases completas estão anonimizadas, seguindo as orientações da Lei Geral de Proteção de Dados.

MÉTODOS

Modelo de Regressão Logística

Defina Y_i como um variável aleatória binária, que assume o valor 1 se o i -ésimo aluno egresso se matriculou em uma pós graduação, e 0 caso contrário, $i = 1, 2, \dots, n$, sendo n o número total de egressos que compõem a amostra em estudo. Assuma que Y_i segue uma distribuição de Bernoulli com parâmetro p_i que representa a probabilidade do i -ésimo aluno ingressar na pós-graduação, ou seja, $p_i = P(Y_i = 1)$. Uma classe de modelos bastante utilizada para

¹<https://www.conre3.org.br/porta1/formacao-em-estatistica/>

tratar variáveis binárias ou dicotômicas é chamada de *modelos de regressão logística* (GIOLO, 2017; PAULA, 2024). O modelo de regressão logística para a variável Y_i pode ser expresso por

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_q x_q, \quad i = 1, \dots, n,$$

em que $(x_1, \dots, x_q)^t$ representa o vetor das variáveis predictoras do modelo e $(\beta_0, \beta_1, \dots, \beta_q)$ o vetor de parâmetros.

Os parâmetros β_j , $j = 2, \dots, q$ podem ser interpretados como efeitos aditivos no logaritmo da razão de chances, e e^{β_j} como a razão de chances. A probabilidade de um egresso ingressar na pós-graduação pode ser então calculada por

$$p_i = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_q x_q}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_q x_q}}.$$

Modelo Aditivo Generalizado de Localização, Dispersão e Forma

Agora, defina Y_i como a variável aleatória contínua que representa a renda média em salários mínimos do i -ésimo egresso que se inseriu no mercado de trabalho dois anos após a conclusão da graduação, para $i = 1, 2, \dots, n$, sendo n o número total de egressos observados. Neste caso, assume-se que Y_i segue uma distribuição da família *skew t-type 2* (ST2), a qual pertence à classe dos modelos aditivos generalizados de localização, dispersão e forma (*Generalized Additive Models for Location, Scale and Shape* – GAMLSS) (STASINOPOULOS; RIGBY, 2007). A estrutura geral do modelo GAMLSS pode ser expressa como

$$Y_i \sim f(y_i | \mu_i, \sigma_i, \nu_i, \tau_i),$$

em que, no caso da distribuição ST2, μ_i representa o parâmetro de localização, σ_i representa o parâmetro de dispersão, ν_i é o parâmetro de assimetria, e τ_i representa o parâmetro de curtose. A distribuição ST2 é capaz de acomodar assimetria e caudas pesadas, características frequentemente observadas em distribuições de renda, o que a torna especialmente adequada para modelar a variável Y_i . Assim, o modelo para Y_i pode ser representado por

$$\mu_i = \beta_0 + \beta_1 x_1 + \dots + \beta_q x_q, \quad i = 1, \dots, n,$$

em que $(x_1, \dots, x_q)^t$ representa o vetor de covariáveis associadas ao egresso e $\beta_0, \beta_1, \dots, \beta_q$ são os coeficientes a serem estimados.

Todas as análises estão sendo realizadas utilizando a linguagem de programação R (R CORE TEAM, 2024).

RESULTADOS

Perfil geral dos alunos egressos

A Figura 1, na qual A representa o conjunto dos pós-graduandos e B , o conjunto dos egressos que seguiram para o mercado de trabalho, permite visualizar a distribuição dos grupos que compõem a amostra de alunos egressos que está sendo analisada neste estudo. Dos 463 estudantes que se formaram em Estatística na Unicamp entre 2005 e 2016, 385 estavam no mercado de trabalho após dois anos e 94 tinham ingressado em uma pós-graduação *stricto sensu* até 2022. Além disso, 53 estudantes estavam em ambos os cenários e 37 em nenhum deles.

A análise desses 463 alunos egressos mostra que eles ingressaram na graduação entre 1997 e 2013, sendo que a maioria (80%) iniciaram entre 2001 e 2009. Estatística era a primeira opção de curso para 74,3% deles. Quanto as notas padronizadas obtidas no vestibular observou-se que: em Português, 23,6% tiraram abaixo de 450, 48,8% entre 450 e 550, e 27,5% acima de 550 pontos; em Matemática, 24,4% tiraram abaixo de 490, 50,1% entre 490 e 559, e 25,5% acima de 559 pontos. A quantidade de homens e mulheres é semelhante, com 227 homens (49%) e 236 mulheres

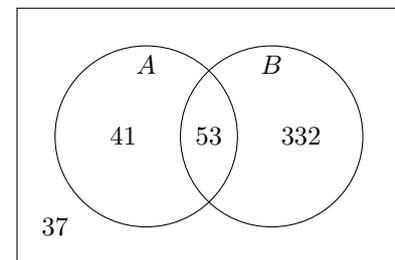


Figura 1: Diagrama de Venn dos grupos que compõem a amostra. O conjunto A representa os egressos com matrícula na pós-graduação após se formarem em Estatística e o conjunto B representa os egressos com carteira assinada após dois anos de formados.

(51%), uma proporção que se mantém estável tanto por ano de ingresso quanto de egresso. Em relação à etnia, 58,1% dos alunos são brancos, enquanto apenas 5,4% se identificam como pretos ou pardos – essa informação, no entanto, passou a ser coletada apenas a partir de 2003 e, portanto, há uma alta porcentagem de registros na categoria “Sem informação”. A maioria dos estudantes (82,3%) cursou o ensino médio comum e predominantemente (68,5%) em escolas particulares. Além disso, mais de 65% dos pais e mães possuem ensino médio ou ensino superior como nível de escolaridade. Quanto à renda, a maioria dos alunos provém de famílias com rendimentos de 5 a 10 salários mínimos ou superiores a 15 salários mínimos. Adicionalmente, 44,3% dos estudantes são não-PAAIS, 11,9% são PAAIS e não há informações sobre o restante, já que esses dados também estão disponíveis apenas a partir de 2005, quando surgiu o Programa de Ação Afirmativa e Social. No momento do egresso, 47,7% tinham entre 21 e 23 anos, 42,5%, entre 24 e 26, e apenas 9,7% tinham mais de 27. O tempo médio para a conclusão do curso entre esses egressos foi de 10,7 semestres, sendo que quase 30% levaram exatamente 10 semestres para se formar e apenas 13,6% concluíram no tempo previsto de 8 semestres. Além disso, o coeficiente de rendimento padrão (CRP) médio entre eles é de 0,62. Vele mencionar também que a renda média, dois anos após a conclusão do curso, foi de 7,9 salários mínimos entre os egressos que se inseriram no mercado de trabalho.

Chance de um aluno concluinte ingressar em uma pós-graduação

A Tabela 1 apresenta os resultados do modelo de regressão logística ajustado para investigar os fatores associados ao ingresso em uma pós-graduação. A variável resposta é uma variável binária que assume o valor 1 quando o aluno ingressou na pós-graduação e 0 caso contrário. Para estimar os efeitos das variáveis explicativas, partiu-se de um modelo inicial mais completo, ajustado com todas as variáveis disponíveis, e, em seguida, as menos significativas foram removidas uma a uma, com base em seus p-valores.

Os resultados indicam que o tempo de conclusão do curso de graduação está negativamente associado à probabilidade de ingresso na pós-graduação. Alunos que concluíram o curso entre 10 e 12 semestres apresentaram uma chance 78,9% menor de ingressar na pós-graduação em comparação àqueles que concluíram em 9 semestres ou menos. Essa redução é ainda mais acentuada entre os alunos que levaram 13 semestres ou mais, cuja chance foi 84,6% menor, mantendo-se constantes as demais variáveis do modelo.

Em relação ao sexo, os resultados mostram que mulheres apresentaram uma chance 49,8% menor de ingressar em uma pós-graduação em comparação aos homens.

O desempenho na prova de língua portuguesa no vestibular também se mostrou um importante preditor. Alunos com nota superior a 550 tiveram uma chance 185,9% maior de ingresso na pós-graduação em relação àqueles com nota inferior a 450. Embora a chance também aumente (em 67,9%) para os alunos com notas entre 450 e 550, essa diferença não foi estatisticamente significativa ao nível de 5%.

Por fim, o vínculo com o mercado de trabalho formal apresentou associação negativa com o ingresso na pós-graduação. Alunos com vínculo CLT logo após a formação tiveram uma chance 84,2% menor de seguir para a pós-graduação, o que pode indicar que a inserção no mercado de trabalho reduz a disposição ou a necessidade de continuidade na formação acadêmica.

Tabela 1: Estimativas dos coeficientes do modelo de regressão logística ajustado.

Variável	Categoria	Estimativa	Erro Padrão	Z	p-valor	e^β	$(e^\beta - 1) \times 100$
(Intercepto)		0,84	0,39	2,17	0,0301	2,31	+131,2%
Tempo de conclusão	9 ou menos (ref.)						
	Entre 10 e 12	-1,56	0,29	-5,45	0,0000	0,21	-78,9%
	13 ou mais	-1,87	0,49	-3,81	0,0001	0,15	-84,6%
Sexo	Masculino (ref.)						
	Feminino	-0,69	0,28	-2,45	0,0142	0,50	-49,8%
Nota de português	< 450 (ref.)						
	≥ 450 e ≤ 550	0,52	0,36	1,44	0,1497	1,68	+67,9%
	> 550	1,05	0,40	2,64	0,0084	2,86	+185,9%
Vínculo empregatício	Não (ref.)						
	Sim	-1,85	0,30	-6,07	0,0000	0,16	-84,2%

O teste de razão de verossimilhança indicou que o modelo final não difere significativamente do modelo inicial que incluía todos os efeitos principais (p -valor = 0,6980), sugerindo que a simplificação do modelo não comprometeu seu ajuste. Além disso, a análise dos resíduos, realizada por meio do gráfico quantil-quantil com envelope simulado, não apresentou nenhum indício de que a distribuição utilizada seja inadequada. Como é possível observar na Figura 2, a maioria dos pontos permaneceu dentro dos limites de confiança simulados, indicando que o modelo final apresenta um ajuste satisfatório aos dados.

Salário médio de um egresso com vínculo CLT

A Tabela 2 apresenta os resultados do modelo GAMLSS ajustado para explicar a renda média, em salários mínimos, dos egressos dois anos após a conclusão do curso, considerando apenas aqueles que ingressaram no mercado de trabalho. Para estimar os efeitos das variáveis explicativas, partiu-se de um modelo inicial mais completo, ajustado com todas as variáveis disponíveis, e, em seguida, as menos significativas foram removidas uma a uma, com base em seus p -valores.

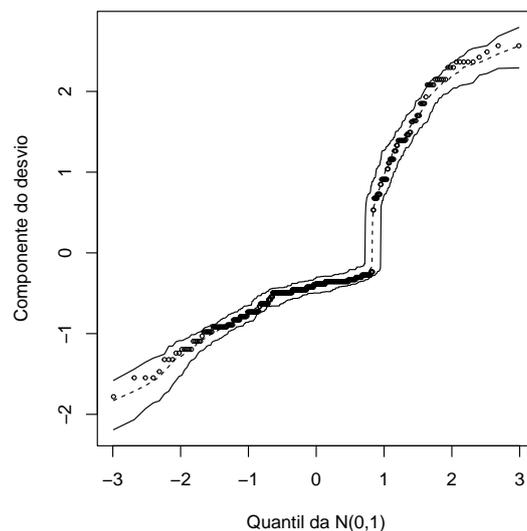


Figura 2: Gráfico normal de probabilidades para o resíduo componente do desvio referente ao modelo logístico ajustado.

Tabela 2: Estimativas dos coeficientes do modelo GAMLSS ajustado.

Variável	Categoria	Estimativa	Erro Padrão	t	p-valor
Intercepto		5,34	0,75	7,12	0,0000
Tempo de conclusão do curso	9 ou menos (referência)				
	10 a 12	-0,53	0,34	-1,55	0,1213
	13 ou mais	-1,46	0,46	-3,19	0,0016
Sexo	Masculino (referência)				
	Feminino	-0,81	0,27	-2,95	0,0034
Renda familiar	Até 3 (referência)				
	3 a 5	0,03	0,68	0,05	0,9600
	5 a 10	0,28	0,61	0,46	0,6493
	10 a 15	0,79	0,65	1,22	0,2232
	Mais de 15	1,48	0,61	2,41	0,0166

Estimativas dos parâmetros adicionais do modelo: $\hat{\sigma} = 1,17$ (erro padrão = 0,12), $\hat{\nu} = 1,88$ (erro padrão = 0,55), $\hat{\tau} = 1,46$ (erro padrão = 0,22)

Os resultados indicam que o tempo de conclusão do curso, o sexo e a renda familiar estão significativamente associados à renda média dos egressos. Alunos que levaram 13 semestres ou mais para concluir o curso apresentaram, em média, uma renda 1,46 salários mínimos inferior àqueles que finalizaram em 9 semestres ou menos.

Além disso, observou-se que mulheres apresentam, em média, uma renda 0,81 salários mínimos menor do que homens dois anos após a formatura, o que evidencia uma desigualdade de gênero nos rendimentos entre egressos do curso de Estatística.

Outro fator relevante foi a renda familiar: alunos oriundos de famílias com renda mensal superior a 15 salários mínimos apresentaram, em média, uma renda 1,48 salários mínimos superior à dos egressos de famílias com até 3 salários mínimos. Isso indica que o contexto socioeconômico de origem exerce influência persistente sobre os rendimentos mesmo após a graduação.

As demais categorias das variáveis incluídas no modelo – como conclusão entre 10 e 12 semestres e faixas intermediárias de renda familiar – não apresentaram efeitos estatisticamente significativos ao nível de 5%.

As estimativas para os parâmetros σ , ν e τ ($\hat{\sigma} = e^{1,17} \approx 3,23$, $\hat{\nu} = 1,88$ e $\hat{\tau} = e^{1,46} \approx 4,29$) indicam que a distribuição da renda apresenta uma dispersão considerável, assimetria à direita ($\nu > 0$) e caudas mais pesadas que a normal ($\tau > 3$). Além disso, todos os parâmetros foram estatisticamente significativos, o que confirma a necessidade

da modelagem completa da forma da distribuição e mostra que o GAMLSS com distribuição ST2 é adequado para capturar as principais características da renda na amostra analisada.

O teste de razão de verossimilhança indicou que o modelo final não difere significativamente do modelo inicial que incluía todos os efeitos principais (p -valor = 0,2228), sugerindo que a simplificação do modelo não comprometeu seu ajuste. Além disso, a análise dos resíduos quantis normalizados, por meio do gráfico quantil-quantil (Figura 3), indica que esses resíduos seguem aproximadamente uma distribuição normal padrão, o que sugere que o modelo está bem ajustado aos dados.

CONCLUSÃO

Em síntese, os resultados obtidos permitem traçar um panorama das trajetórias dos egressos do curso de Estatística da Unicamp, evidenciando desigualdades tanto no ingresso em programas de pós-graduação quanto na inserção e rendimento no mercado de trabalho. Fatores como tempo de conclusão do curso, desempenho no vestibular, sexo e renda familiar mostraram-se significativamente associados a esses desfechos, indicando que, mesmo após a conclusão do ensino superior, persistem efeitos estruturais que moldam as oportunidades acadêmicas e profissionais desses egressos.

BIBLIOGRAFIA

- GIOLO, S. R. **Introdução à análise de dados categóricos com aplicações**. São Paulo: Blucher, 2017.
- PAULA, G. A. **Modelos de regressão com apoio computacional**. São Paulo: IME-USP, 2024.
- R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria: R Foundation for Statistical Computing, 2024. <https://www.R-project.org>.
- RIBEIRO, C. A. C.; SCHLEGEL, R. Estratificação horizontal da educação superior no Brasil (1960 a 2010). In: ARRETCHE, Marta (Ed.). **Trajetoárias das desigualdades: como o Brasil mudou nos últimos cinquenta anos**. São Paulo: Editora Unesp, 2015.
- SAITO, C. Y. **As desigualdades nos retornos do ensino superior no Brasil**. [S.l.: s.n.], 2006. https://www.ipea.gov.br/ipeacaixa/premio2006/docs/trabpremiados/IpeaCaixa2006_Estudante_MH01_tema02.pdf. Prêmio IPEA-CAIXA.
- SIMON, L.; PACHECO, A. Ações de acompanhamento de egressos: um estudo das universidades públicas do sul do Brasil. **Revista Brasileira de Ensino Superior**, Passo Fundo, v. 3, n. 2, p. 94–113, 2017. DOI: 10.18256/2447-3944.2017.v3i2.2023.
- STASINOPOULOS, D. M.; RIGBY, R. A. Generalized Additive Models for Location, Scale and Shape (GAMLSS) in R. **Journal of Statistical Software**, v. 23, n. 7, dez. 2007. Disponível em: <https://www.jstatsoft.org/v23/i07/i>.
- VAZ, K. C. d. S. **Trajetoárias ocupacionais dos egressos do Ensino Superior**. 2022. Tese (Doutorado em Sociologia) – Universidade de Brasília, Brasília.

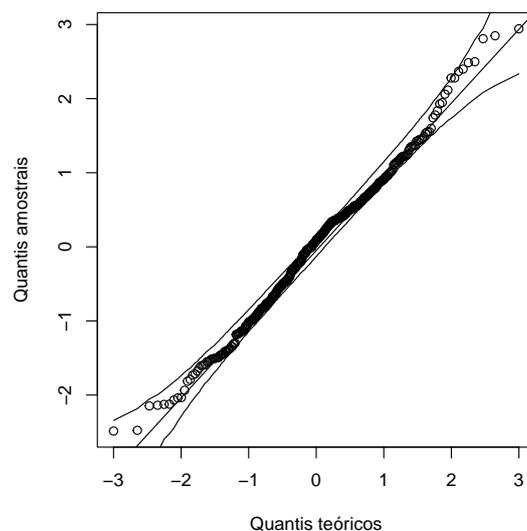


Figura 3: Gráfico quantil-quantil dos resíduos referentes ao modelo GAMLSS ajustado.