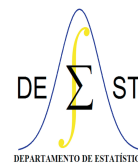




UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
DEPARTAMENTO DE ESTATÍSTICA



DOUGLAS OLIVEIRA GOMES

ANÁLISE ESTATÍSTICA DO EFEITO DA PARTICIPAÇÃO
EM PROGRAMAS DE INICIAÇÃO CIENTÍFICA NO
INGRESSO EM PROGRAMAS DE PÓS-GRADUAÇÃO NA
UNIVERSIDADE FEDERAL DE VIÇOSA

DOUGLAS OLIVEIRA GOMES

ANÁLISE ESTATÍSTICA DO EFEITO DA PARTICIPAÇÃO EM
PROGRAMAS DE INICIAÇÃO CIENTÍFICA NO INGRESSO EM
PROGRAMAS DE PÓS-GRADUAÇÃO NA UNIVERSIDADE FEDERAL
DE VIÇOSA

Trabalho de conclusão de curso apresentado
ao curso de Estatística da Universidade Fe-
deral de Ouro Preto como requisito parcial
para a obtenção do título de Bacharel em
Estatística.

Orientador: Helgem de Souza Ribeiro Mar-
tins

OURO PRETO

2026

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

G633a Gomes, Douglas Oliveira.

Análise estatística do efeito da participação em programas de iniciação científica no ingresso em programas de pós-graduação na Universidade Federal de Viçosa. [manuscrito] / Douglas Oliveira Gomes. - 2026.

30 f.: il.: gráf., tab..

Orientador: Prof. Dr. Helgem de Souza Ribeiro Martins.

Monografia (Bacharelado). Universidade Federal de Ouro Preto. Instituto de Ciências Exatas e Biológicas. Graduação em Estatística .

1. Pós-graduação. 2. Pesquisa. 3. Regressão Logística. 4. Análise Estatística. I. Martins, Helgem de Souza Ribeiro. II. Universidade Federal de Ouro Preto. III. Título.

CDU 519.23:001.891

Bibliotecário(a) Responsável: Renata Mara de Almeida - CRB-7: 6328



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE OURO PRETO
REITORIA
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
COLEGIADO DO CURSO DE ESTATÍSTICA



FOLHA DE APROVAÇÃO

Douglas Oliveira Gomes

Análise estatística do efeito da participação em programas de iniciação científica no ingresso em programas de pós-graduação na Universidade Federal de Viçosa

Monografia apresentada ao Curso de Estatística da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Estatística

Aprovada em 27 de janeiro de 2026

Membros da banca

Dr. Helgem de Souza Ribeiro Martins - Orientador (Universidade Federal de Ouro Preto)
Dr. Eduardo Bearzoti (Universidade Federal de Ouro Preto)
Dr. Eduardo Campana Barboza (Universidade Federal de Viçosa)

Professor Dr. Helgem de Souza Ribeiro Martins, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 27/01/2026



Documento assinado eletronicamente por **Helgem de Souza Ribeiro Martins, PROFESSOR DE MAGISTERIO SUPERIOR**, em 27/01/2026, às 16:52, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Eduardo Bearzoti, PROFESSOR DE MAGISTERIO SUPERIOR**, em 28/01/2026, às 12:00, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1048458** e o código CRC **11C0BFF9**.

Agradecimentos

Gostaria de expressar, primeiramente, minha profunda gratidão a Deus, por sempre me abençoar e me conceder forças ao longo de toda esta trajetória. Durante muito tempo, a conclusão deste curso pareceu algo distante da minha realidade, chegando, inclusive, a duvidar de mim mesmo em diversos momentos. No entanto, Deus, em Sua infinita misericórdia, mostrou-me de forma incansável que esse sonho era possível e que, com perseverança, eu seria capaz de alcançá-lo.

Agradeço imensamente à minha mãe, que sempre depositou confiança em mim e foi fundamental na formação do meu caráter. Por meio de muita luta, dedicação e resiliência diante das adversidades, ela me ensinou a ser uma pessoa melhor e a não desistir diante das dificuldades. Ao meu irmão, que esteve ao meu lado durante toda essa caminhada, oferecendo apoio constante e jamais deixando de acreditar em mim. Estendo também meus agradecimentos à minha avó e ao meu avô, que, mesmo diante de um contexto familiar com pouco acesso aos estudos, sempre me incentivaram e apoiaram em algo completamente novo para nossa realidade.

Aos meus irmãos, deixo registrado meu sincero reconhecimento. Ao meu irmão Iago, agradeço pelos valiosos conselhos, pelo apoio contínuo e pela confiança em meu potencial. Ao meu irmão Maycon, que, mesmo à distância, sempre esteve presente por meio de palavras de incentivo, renovando minhas energias nos momentos mais desafiadores. Ao meu irmão Kaiky, agradeço por tornar os dias mais leves com sua alegria contagiante e por estar sempre disposto a ajudar quando precisei. Ao meu irmão Alexandre, que também é companheiro de curso, agradeço por compartilhar conhecimentos, conselhos e bons momentos, além de dividir comigo os desafios, obstáculos e perrengues vividos em Ouro Preto, tornando essa jornada mais suportável e enriquecedora.

Registro minha sincera gratidão ao meu professor e orientador, Helgem, por todos os ensinamentos, orientações e conselhos ao longo deste trabalho, bem como pela confiança depositada em meu potencial. Sou igualmente grato a todos os professores e técnicos do Departamento de Estatística (DEEST), que, com dedicação e compromisso, contribuíram de forma decisiva para a concretização deste sonho.

Por fim, agradeço à Universidade Federal de Ouro Preto (UFOP) e às políticas de permanência e auxílio estudantil, que foram essenciais para viabilizar a conclusão deste curso e a obtenção deste diploma. Estendo meus agradecimentos a todos os funcionários da instituição, que desempenham suas funções com excelência e contribuem diariamente para a formação de inúmeros estudantes.

“A educação é a arma mais poderosa que você pode usar para mudar o mundo.”
(Nelson Mandela)

Resumo

Este estudo investiga a relação entre a participação em Iniciação Científica (IC) e as taxas de ingresso e sucesso em programas de pós-graduação da Universidade Federal de Viçosa (UFV). A pesquisa foi motivada pela necessidade de compreender os fatores que influenciam o acesso e a permanência acadêmica, especialmente no contexto brasileiro, em que as taxas de evasão na pós-graduação representam um desafio tanto para os estudantes quanto para as instituições. Utilizando técnicas de estatística descritiva e regressão logística, os dados foram obtidos por meio do portal de dados abertos da Universidade Federal de Viçosa, e análises específicas foram realizadas com o software R. Variáveis como idade, gênero, estado de origem e situação acadêmica foram examinadas, possibilitando a identificação de padrões e tendências nas trajetórias dos estudantes. Foi possível traçar um perfil médio dos estudantes da UFV, tanto na graduação quanto na pós-graduação, e avaliar a relação entre a iniciação científica e os estudos de pós-graduação. Os resultados indicam que a participação em iniciação científica apresenta associação positiva com o ingresso na pós-graduação. Estudantes que participaram da iniciação científica durante a graduação apresentaram uma taxa significativamente maior de continuidade acadêmica em comparação àqueles que não participaram desse programa. Os achados desta pesquisa ressaltam a importância de políticas institucionais que incentivem a iniciação científica, promovendo maior engajamento acadêmico e contribuindo para a redução das taxas de evasão nos programas de pós-graduação. O estudo também contribui para a base de conhecimento sobre a relação entre a iniciação científica e a permanência em programas de pós-graduação.

Palavras-chave: estudos de pós-graduação; iniciação científica; ciência de dados; regressão logística; evasão escolar.

Abstract

This study investigates the relationship between participation in Scientific Initiation (SI) and admission and success rates in graduate programs at the Federal University of Viçosa (UFV). The research was motivated by the need to understand the factors that influence access and academic continuity, especially in the Brazilian context, where graduate school dropout rates pose a challenge for both students and institutions. Using descriptive statistics and logistic regression techniques, data were obtained through the Federal University of Viçosa's open data portal, and specific data were analyzed using R software. Variables such as age, gender, state of origin, and academic standing were examined, enabling the identification of patterns and trends in student trajectories. It was possible to draw an average profile of UFV students, both at the undergraduate and graduate levels, and to assess the relationship between scientific initiation and graduate studies. The results indicate that participation in scientific initiation has a positive association with admission to graduate studies. Students who participated in scientific initiation during their undergraduate years had a significantly higher rate of academic continuation compared to those who did not participate in this program. The results of this research underscore the importance of institutional policies that encourage scientific initiation, promoting greater academic engagement and contributing to a reduction in dropout rates in graduate programs. The study also contributes to the knowledge base on the relationship between scientific initiation and retention in graduate programs.

Keywords: postgraduate studies; scientific initiation; data science; logistic regression; school dropout

Lista de ilustrações

Figura 1 – Distribuição dos estudantes da UFV por sexo e nível de formação . .	18
Figura 2 – Distribuição dos estudantes da UFV por nacionalidade e nível de formação	19
Figura 3 – Distribuição dos estudantes da UFV por faixa etária e nível de formação	20
Figura 4 – Distribuição dos estudantes da UFV por área de estudo e nível de formação	20
Figura 5 – Distribuição dos estudantes da UFV por raça e nível de formação . .	21
Figura 6 – Distribuição dos estudantes da UFV por situação acadêmica e nível de formação	22
Figura 7 – Distribuição dos estudantes da UFV por UF e nível de formação . . .	23
Figura 8 – Distribuição dos estudantes da UFV por situação acadêmica, nível de formação e IC	24
Figura 9 – Gráfico de envelope: ajuste dos resíduos do modelo de ingresso na pós-graduação	26
Figura 10 – Gráfico de envelope: ajuste dos resíduos do modelo de conclusão da pós-graduação	27

Lista de tabelas

Tabela 1 – Variáveis componentes da base de dados dos bolsistas de IC.	13
Tabela 2 – Variáveis componentes da base de dados dos alunos ingressantes da UFV.	14
Tabela 3 – Variáveis criadas no banco de dados.	15
Tabela 4 – Associação entre participação em iniciação científica, ingresso e conclusão da pós-graduação	24
Tabela 5 – Estimativas do modelo de regressão logística para o ingresso na pós-graduação segundo participação em iniciação científica	25
Tabela 6 – Estimativas do modelo de regressão logística para a conclusão da pós-graduação segundo participação em iniciação científica	26

Sumário

1	INTRODUÇÃO	10
1.1	Referencial Teórico	11
1.2	Objetivos	11
1.3	Estrutura do Trabalho	12
2	ASPECTOS METODOLÓGICOS	13
2.1	Descrição e construção da base de dados	13
2.2	Metodologia aplicada	15
2.2.1	Razão de chances (<i>odds ratio</i>)	16
2.2.2	Avaliação da adequação do modelo	17
2.2.3	Ambiente de análise e recursos computacionais	17
3	RESULTADOS ALCANÇADOS	18
3.1	Caracterização da amostra	18
3.2	Regressão Logística — Ingresso na pós-graduação	25
3.2.1	Avaliação do ajuste do modelo de ingresso na pós-Graduação	25
3.3	Regressão Logística — Conclusão da pós-graduação	26
3.3.1	Avaliação do ajuste do modelo de conclusão da pós-graduação	27
3.4	Discussão	28
3.5	Limitações do Estudo	28
4	CONSIDERAÇÕES FINAIS	29
	REFERÊNCIAS	30

1 Introdução

A Iniciação Científica (IC) consiste em um instrumento de formação do discente de graduação, com o objetivo de introduzir o aluno no universo da pesquisa científica. Segundo [Bastos *et al.* 2010], a iniciação científica constitui um instrumento de apoio teórico e metodológico que possibilita a realização de um projeto que contribui diretamente para a formação profissional do aluno.

Além da formação científica, a Iniciação Científica no Brasil fornece recursos financeiros aos alunos que participam da modalidade. Segundo [Massi e Queiroz 2010], o financiamento de pesquisas científicas em nível de graduação passou a ser oferecido na década de 1950, com a criação do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), responsável pela concessão de bolsas de fomento à pesquisa na graduação.

Atualmente, o financiamento dessas pesquisas se origina de diversas fontes, incluindo o CNPq, fundações de amparo à pesquisa (FAPs), organizações não governamentais, institutos sem fins lucrativos, empresas e as próprias instituições de ensino superior, com destaque para as Instituições Federais de Ensino Superior (IFES), que contam com financiamento interno para bolsas de iniciação científica. É comum que discentes participantes desenvolvam interesse aprofundado pela ciência, optando por prosseguir sua formação em nível de pós-graduação, fenômeno considerado natural na trajetória acadêmica.

Sob a ótica da gestão acadêmica, a manutenção e a ampliação dessas bolsas são fundamentais para a eficácia das políticas institucionais. De acordo com [Gorgens 2007], a avaliação da gestão em instituições federais deve pautar-se em indicadores que mensurem a qualidade e o impacto social das ações implementadas. Nesse sentido, o fomento à pesquisa na graduação deixa de ser apenas um auxílio financeiro, tornando-se um indicador de desempenho institucional, uma vez que a bolsa viabiliza a dedicação necessária para que o discente atinja os níveis de excelência exigidos tanto na graduação quanto na futura inserção em programas de pós-graduação.

A Universidade Federal de Viçosa (UFV) oferece diversos programas de iniciação científica, incluindo modalidades com fomento externo e participação voluntária, além de um amplo ecossistema de programas de pós-graduação *stricto sensu*. Observa-se, portanto, grande interação entre a IC e a pós-graduação na instituição, uma vez que muitos alunos de IC participam de equipes de pesquisa de mestrado e doutorado.

Diante desse cenário, avaliar o nível de interação entre essas atividades, bem como mensurar o impacto da IC sobre o ingresso e o desempenho de alunos de pós-

graduação, configura-se como uma demanda institucional relevante.

1.1 Referencial Teórico

A prática de iniciar alunos de graduação em projetos científicos remonta ao século XIX. Universidades alemãs incorporaram a pesquisa científica ao processo de formação, buscando desenvolver profissionais e professores mais capacitados e dotados de pensamento científico. A partir desse modelo, consolidou-se a prática contemporânea, em que o estudante é orientado por um supervisor versado na metodologia científica e em uma área específica do conhecimento [Lordelo e Argôlo 2015].

Durante a realização dos projetos de IC, os alunos são familiarizados com aspectos teóricos da metodologia científica, incluindo o processo científico, a redação de textos científicos e a condução de pesquisas. Paralelamente, são introduzidos às metodologias específicas de sua área de estudo, o que favorece seu desenvolvimento enquanto cientistas e futuros profissionais.

Diversos estudos investigaram a relação entre a experiência em IC e a trajetória acadêmica posterior. [Lordelo e Argôlo 2015] analisaram dados de 895 estudantes de mestrado e concluíram que a IC acelera a transição para a Pós-graduação. De modo semelhante, [Pelinski, Kampa *et al.* 2021], utilizando estatísticas descritivas e *Propensity Score Matching* (PSM), demonstrou que a IC exerce impacto significativo nas taxas de admissão, especialmente nas áreas de ciências exatas, biológicas e sociais aplicadas.

Adicionalmente, [Lima 2016] revelou que alunos com experiência prévia em pesquisa se titulam mais jovens, em menor tempo e apresentam maior produção científica, evidenciando a relevância institucional do tema.

1.2 Objetivos

O objetivo geral deste trabalho é analisar a associação entre a participação em programas de Iniciação Científica e a trajetória acadêmica dos estudantes da UFV, com ênfase no ingresso e no desempenho em cursos de pós-graduação *stricto sensu*.

Para alcançar este propósito, os objetivos específicos são:

- Efetuar o levantamento e tratamento de dados institucionais relacionados à trajetória dos discentes;
- Caracterizar o perfil sociodemográfico e acadêmico dos alunos de IC e de pós-graduação;

- Verificar o percentual de alunos de IC que optam pela continuidade dos estudos na instituição;
- Avaliar, por meio de modelos de regressão logística, o impacto da IC na probabilidade de admissão e conclusão da pós-graduação;
- Propor uma abordagem metodológica baseada em dados que auxilie no monitoramento da trajetória acadêmica e no fortalecimento de políticas institucionais.

1.3 Estrutura do Trabalho

Este trabalho encontra-se estruturado em capítulos que apresentam os procedimentos adotados e a interpretação dos resultados:

O Capítulo 2, intitulado Aspectos Metodológicos, descreve detalhadamente os métodos empregados, as bases de dados e as técnicas estatísticas aplicadas. No Capítulo 3, Resultados Alcançados, são apresentadas as análises descritivas e inferenciais, incluindo modelos de regressão logística, representações gráficas e uma breve discussão que se dedica à análise crítica dos achados em diálogo com a literatura. No Capítulo 4, Considerações Finais, sintetizam-se as principais evidências e contribuições para o ambiente acadêmico. Por fim, as Referências listam as obras consultadas e citadas ao longo deste trabalho.

2 Aspectos Metodológicos

O desenvolvimento deste trabalho envolveu a análise de dados de iniciação científica e pós-graduação, com o objetivo de compreender padrões, características e indicadores relacionados à produção acadêmica e ao perfil dos pesquisadores. Para isso, tornou-se essencial utilizar uma base de dados acessível, consistente e suficientemente detalhada para permitir análises estatísticas confiáveis.

Diante das restrições de acesso a determinadas bases institucionais, optou-se por utilizar os dados disponibilizados pela Universidade Federal de Viçosa em seu portal de dados abertos. Essa escolha permitiu a construção de um conjunto amplo de informações, favorecendo a aplicação das técnicas analíticas propostas e garantindo a transparência e a reprodutibilidade do estudo.

É importante destacar que as metodologias adotadas, tanto os procedimentos de tratamento dos dados quanto as análises estatísticas realizadas, são plenamente aplicáveis a outras instituições de ensino superior.

2.1 Descrição e construção da base de dados

Os dados utilizados nesta pesquisa foram obtidos diretamente do portal de dados abertos da Universidade Federal de Viçosa, e podem ser acessados por meio do link <https://dados.ufv.br> e organizados para análise. A primeira base de dados em estudo é composta por 7 variáveis e 20.825 observações de participação em programas de Iniciação Científica (IC) na UFV, abrangendo o período de 1990 a 2025. Inclui variáveis como identificador do estudante, curso, título do projeto, orientador, tipo de bolsa e datas de início e término.

Tabela 1 – Variáveis componentes da base de dados dos bolsistas de IC.

Variável	Descrição
id_unico_estudante	Identificador único do estudante.
curso_aluno	Curso que o aluno está matriculado.
titulo_projeto	Título do projeto de Iniciação Científica.
orientador	Nome do professor orientador do projeto.
tipo_bolsa	Tipo de bolsa recebida pelo estudante.
data_inicio	Data de início do projeto de IC.
data_fim	Data de término do projeto de IC.

A segunda base de dados a ser analisada contém 23 variáveis e 96.091 observações sobre os estudantes da Universidade Federal de Viçosa (UFV), abrangendo o período de 2010 a 2024. As variáveis registradas incluem dados demográficos, como sexo, ano de nascimento, raça, naturalidade, estado de origem, e país, além de informações sobre a forma de ingresso na universidade (vestibular, portador de diploma, entre outros) e se o estudante pertence a algum grupo de cota. Também são apresentados dados acadêmicos, como o curso de graduação, nível agrupado, área do curso, turno, situação do aluno (conclusão, evasão, desligamento) e o campus frequentado.

Tabela 2 – Variáveis componentes da base de dados dos alunos ingressantes da UFV.

Variável	Descrição
CodigoEstudante	Identificador único do estudante.
Sexo	Variável categórica que indica o gênero do aluno.
AnoNascimento	Ano de nascimento do estudante.
Raca	Raça/Cor autodeclarada.
Naturalidade	Cidade natal do estudante.
UF	Estado de origem do estudante.
País	País de origem do estudante.
FormaIngresso	Forma de ingresso do estudante.
GrupoCotaConvocado	Categoria de cota do estudante, se aplicável.
AnoAdmissao	Ano de admissão do estudante na universidade.
Curso	Curso em que o estudante está matriculado.
NivelAgrupado	Nível de agrupamento do curso (Graduação ou Pós-Graduação).
AreaCurso	Classificação do curso por área do conhecimento.
Turno	Turno em que o estudante cursou.
SituacaoAluno	Status acadêmico do estudante na universidade.
SituacaoAlunoAgrupada	Classificação geral da situação acadêmica do estudante.
Campus	Campus da universidade onde o estudante está matriculado.

A estratégia adotada consistiu na unificação das bases de dados, por meio da utilização de um identificador único presente em ambas para rastrear quais alunos estavam nas duas fontes. Esse processo permitiu identificar aqueles que participaram de Iniciação Científica (IC) durante a graduação e posteriormente ingressaram na pós-graduação. Como uma das bases possuía um volume significativamente maior de registros em relação à outra, foi estabelecido um critério temporal, filtrando-se a base por meio da variável "AnoAdmissao" para considerar apenas o período entre 2010 e 2024. Essa filtragem foi essencial para tornar a análise mais precisa e gerenciável. Após essa etapa, foi criada uma única base de dados, que reúne as informações relevantes de ambas as fontes, o que facilitou a realização das análises.

Após a consolidação da base de dados final, foram aplicadas diversas transformações e análises para facilitar a interpretação dos dados. Primeiramente, foi criada a variável idade, obtida a partir da subtração do ano de nascimento do ano atual (2024). Em seguida, a variável faixa etária foi gerada para segmentar os alunos em categorias específicas: até 21 anos, 22–25, 26–30, 31–35, 36–40 e 41 anos ou mais.

Além disso, foi criada a variável IC, utilizando o identificador único presente nas duas bases para verificar a participação do aluno na Iniciação Científica. Essa variável recebeu o valor "Sim" para alunos presentes em ambas as bases e "Não" caso contrário. Da mesma forma, a variável pós-graduação foi gerada para indicar se o aluno ingressou na pós-graduação, com valor "Sim" para aqueles identificados na base da pós-graduação e "Não" para os que não ingressaram. Para facilitar as análises, foi incluída a variável graduação, que assume o valor "Sim" para todos os alunos, pois a base foi construída a partir da graduação.

Para aprimorar a análise por estado de origem (UF), os estados com baixa representatividade foram agrupados na categoria "Demais estados", enquanto os registros sem informação foram classificados como "NI" (Não Informado). Por fim, foi criada a variável nacionalidade, diferenciando alunos brasileiros e estrangeiros com base no país de origem.

Tabela 3 – Variáveis criadas no banco de dados.

Variável	Descrição
faixa_idade	Faixa etária do estudante, categorizada.
nacionalidade	Identificação do estudante como Brasileiro ou Estrangeiro.
uf_agrupado	Estado de origem agrupado.
ic	Indicador de participação na Iniciação Científica (Sim/Não).
pos_graduacao	Indicador de ingresso na Pós-Graduação (Sim/Não).

2.2 Metodologia aplicada

Este trabalho utilizou a análise descritiva e a regressão logística como principais ferramentas metodológicas para responder às questões apresentadas na seção de objetivos.

A Estatística Descritiva tem como finalidade organizar, resumir e apresentar os dados de forma a facilitar sua interpretação, por meio de métodos numéricos e gráficos, além de subsidiar análises inferenciais posteriores [Montgomery 2017].

Para investigar a associação entre as características dos estudantes, a participação em programas de Iniciação Científica (IC) e o ingresso na pós-graduação, empregou-se a regressão logística, técnica amplamente utilizada para modelar variáveis resposta

dicotômicas. O modelo busca estimar a probabilidade de ocorrência de um evento de interesse em função de um conjunto de variáveis explicativas [Agresti 2002].

O modelo de regressão logística pode ser expresso como:

$$\log \left(\frac{\pi(x)}{1 - \pi(x)} \right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p, \quad (2.1)$$

onde $\pi(x)$ representa a probabilidade de sucesso associada ao vetor de variáveis explicativas (x_1, x_2, \dots, x_p) , e $(\beta_0, \beta_1, \dots, \beta_p)$ são os parâmetros do modelo.

2.2.1 Razão de chances (*odds ratio*)

Na regressão logística, a probabilidade de sucesso pode ser escrita, de forma equivalente, como:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p}}. \quad (2.2)$$

A chance (*odds*) é definida como a razão entre a probabilidade de ocorrência do evento e a probabilidade de não ocorrência:

$$\text{odds}(x) = \frac{\pi(x)}{1 - \pi(x)}. \quad (2.3)$$

Substituindo a expressão de $\pi(x)$, obtém-se:

$$\text{odds}(x) = e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p}. \quad (2.4)$$

Dessa forma, o modelo de regressão logística estabelece uma relação linear entre as variáveis explicativas e o logaritmo natural da razão de chances. Considerando, para fins de interpretação, uma variação unitária na variável explicativa x_j , mantendo as demais constantes, a razão entre as chances associadas a $x_j + 1$ e x_j é dada por:

$$\frac{\text{odds}(x_j + 1)}{\text{odds}(x_j)} = e^{\beta_j}. \quad (2.5)$$

Essa quantidade é denominada *odds ratio* (OR) e mede o efeito da variável explicativa x_j sobre a chance de ocorrência do evento de interesse. Quando x_j é uma variável quantitativa, a OR indica o fator pelo qual a chance do evento se multiplica a cada aumento unitário em x_j , mantendo as demais variáveis constantes. Quando x_j é uma variável categórica (incluindo o caso dicotômico), a OR representa a razão entre as chances do evento em uma determinada categoria em relação à categoria de referência.

Valores de $OR > 1$ indicam aumento da chance, enquanto valores de $OR < 1$ indicam redução, em relação à categoria ou valor de referência.

Não foi realizado procedimento de seleção de variáveis. Caso a seleção das variáveis explicativas fosse necessária, seria adotado o procedimento *stepwise*, que consiste na inclusão e/ou exclusão sucessiva de variáveis com base em critérios estatísticos, visando à obtenção de um modelo parcimonioso e estatisticamente consistente.

A análise foi conduzida em duas etapas. Inicialmente, realizou-se uma análise descritiva para examinar a distribuição das observações segundo características demográficas e acadêmicas, utilizando tabelas de contingência e representações gráficas, de modo a identificar padrões gerais e possíveis associações preliminares. Em seguida, empregou-se a regressão logística como ferramenta inferencial para quantificar a associação entre as variáveis explicativas e a probabilidade de ocorrência do evento de interesse.

2.2.2 Avaliação da adequação do modelo

A adequação dos modelos de regressão ajustados foi avaliada por meio de gráficos de envelope de simulação, uma ferramenta de diagnóstico amplamente utilizada em modelos de regressão para verificar a consistência entre os resíduos observados e o comportamento esperado sob o modelo teórico.

Esses gráficos são construídos a partir de simulações baseadas no modelo ajustado, gerando uma faixa de variação esperada dos resíduos. Quando a maior parte dos resíduos observados permanece dentro dos limites do envelope, considera-se que o modelo apresenta bom ajuste aos dados. Por outro lado, a presença sistemática de observações fora dessa faixa pode indicar possíveis violações das suposições do modelo, como especificação inadequada, presença de valores atípicos ou inadequação da forma funcional adotada [Warton 2023].

2.2.3 Ambiente de análise e recursos computacionais

As análises foram realizadas no software *R*, com apoio de diferentes pacotes. O pacote *dplyr* foi utilizado para manipulação e transformação dos dados, enquanto *stringr* auxiliou na padronização de textos. A construção dos gráficos foi feita com o *ggplot2*, com formatação de eixos e proporções por meio do pacote *scales*. A validação dos modelos foi conduzida com gráficos de envelope gerados pelo pacote *hnp*. O carregamento otimizado dos pacotes foi realizado utilizando o pacote *pacman*.

3 Resultados Alcançados

Neste capítulo são apresentados os principais resultados obtidos a partir da análise dos dados desta pesquisa. Inicialmente, realiza-se uma caracterização descritiva da amostra, com o objetivo de apresentar o perfil dos estudantes segundo variáveis sociodemográficas, acadêmicas e institucionais. Em seguida, são apresentados os resultados dos modelos de regressão logística, utilizados para avaliar a associação entre a participação em programas de Iniciação Científica e o sucesso acadêmico (ingresso e conclusão da pós-graduação). Os resultados são expostos por meio de tabelas e gráficos, acompanhados de interpretações que auxiliam na compreensão dos achados empíricos.

3.1 Caracterização da amostra

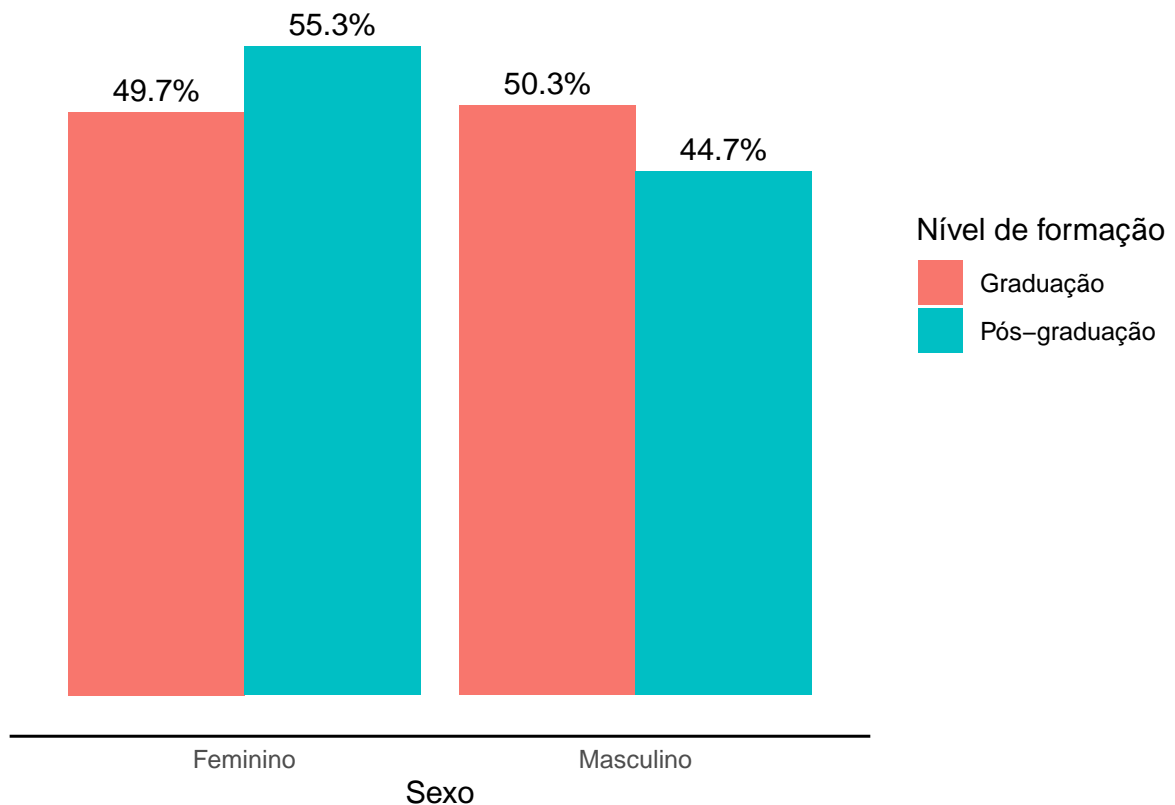


Figura 1 – Distribuição dos estudantes da UFV por sexo e nível de formação

Na Figura 1, a distribuição por sexo mostra-se equilibrada na graduação, onde as mulheres representam 49,7% do corpo discente. Já na pós-graduação, observa-se uma maior presença feminina, alcançando 55,3% dos registros. Entre os homens, a situação se inverte levemente, 50,3% estão na graduação e 44,7% na pós-graduação, revelando menor continuidade dos estudos após a graduação em comparação às mulheres.

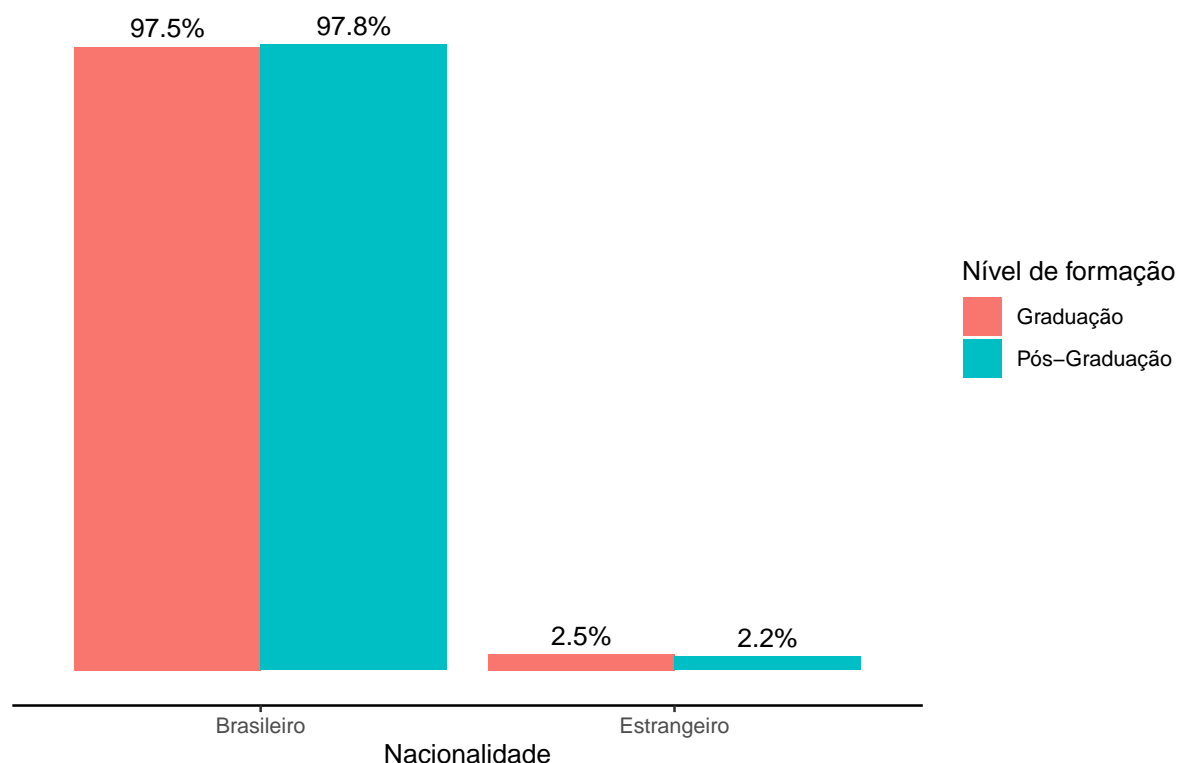


Figura 2 – Distribuição dos estudantes da UFV por nacionalidade e nível de formação

A Figura 2 evidencia que a UFV mantém uma composição discente majoritariamente brasileira, superando 97% tanto na graduação quanto na pós-graduação. Embora a parcela de estudantes estrangeiros seja numericamente reduzida entre 2,2% e 2,5%, esse contingente é institucionalmente significativo. Tal presença ratifica o prestígio internacional da universidade e sua maturidade científica, consolidando o campus como um polo de intercâmbio multicultural que atrai talentos globais e qualifica a produção acadêmica.

Conforme apresentado na Figura 3, os alunos da graduação concentram-se entre 26 e 35 anos, com destaque para as faixas de 26–30 (36,4%) e 31–35 (34,5%). Já na pós-graduação, o maior percentual está na faixa de 31–35 anos (51%), seguida de 26–30 anos (30,6%). E nota-se que, a partir dos 41 anos de idade temos uma descontinuidade dos estudos.

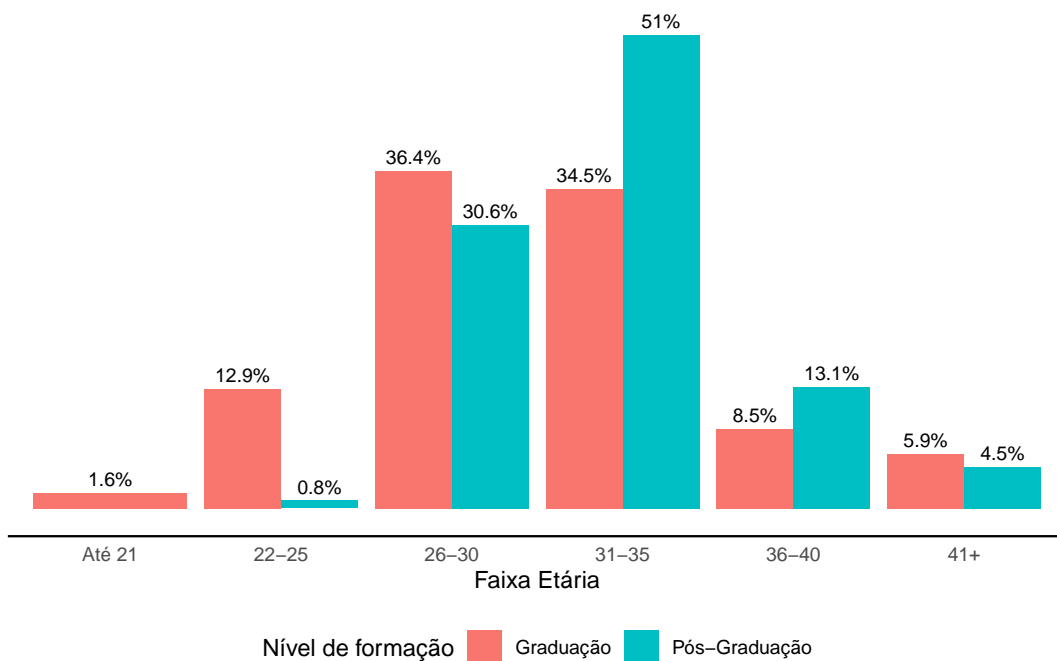


Figura 3 – Distribuição dos estudantes da UFV por faixa etária e nível de formação

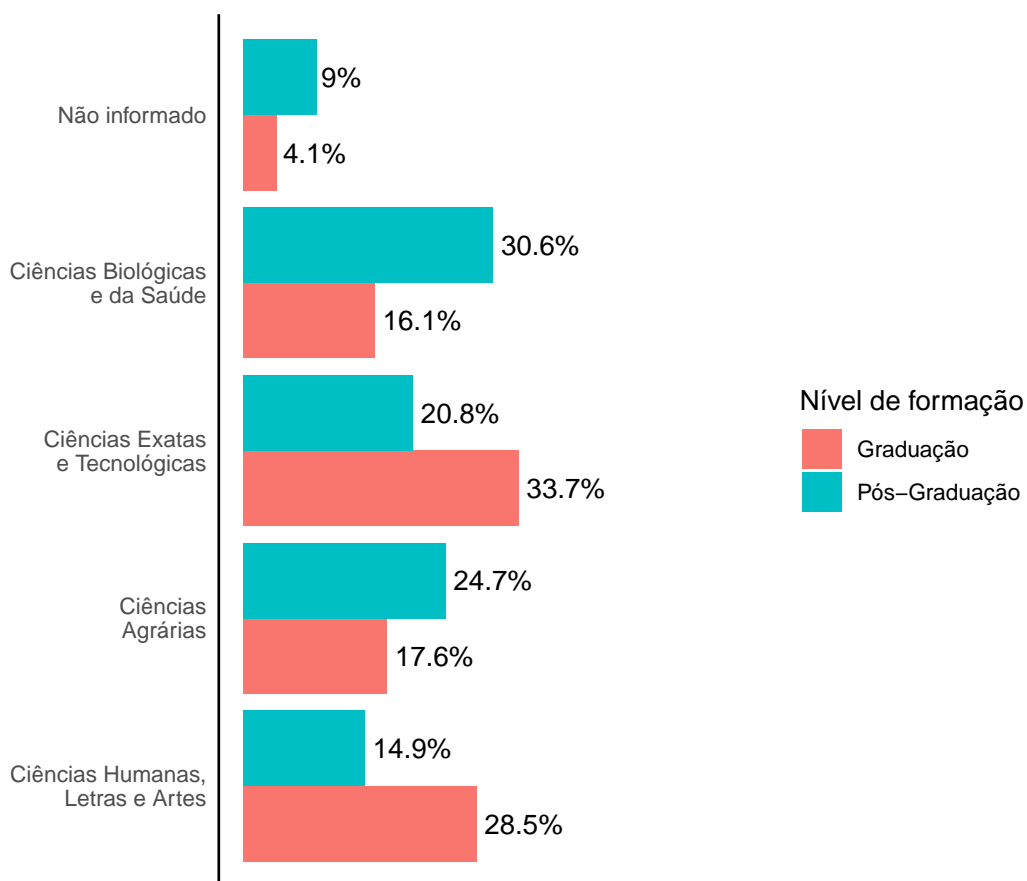


Figura 4 – Distribuição dos estudantes da UFV por área de estudo e nível de formação

A Figura 4 mostra que, na graduação, predominam as Ciências Exatas e Tecnológicas (33,7%) e as Ciências Humanas, Letras e Artes (28,5%). Já na pós-graduação, destacam-se as Ciências Biológicas e da Saúde (30,6%) e as Ciências Agrárias (24,7%). Observa-se, portanto, que as áreas de maior presença na pós-graduação diferem da graduação, com ênfase em campos ligados à saúde e ciências agrárias.

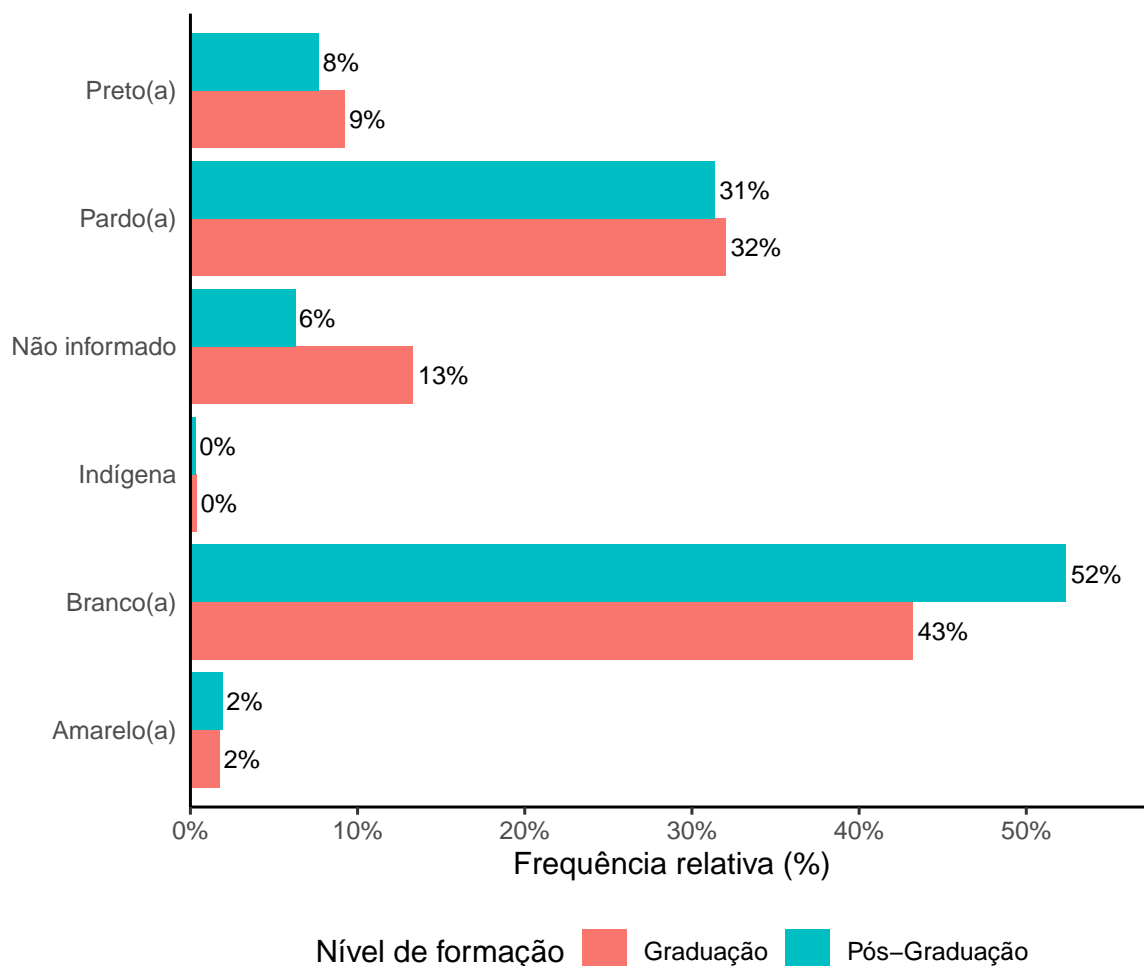


Figura 5 – Distribuição dos estudantes da UFV por raça e nível de formação

A Figura 5 indica que, em ambos os níveis de formação, predominam estudantes autodeclarados brancos, com proporção ligeiramente maior na pós-graduação. As categorias preto e pardo apresentam distribuições semelhantes entre graduação e pós-graduação, sem diferenças expressivas. As categorias amarela e indígena mantêm frequências residuais, e os registros classificados como “não informado” permanecem relevantes, o que pode refletir limitações no preenchimento dos dados. De forma geral, o gráfico sugere a persistência de desigualdades raciais ao longo da trajetória acadêmica, com maior concentração de estudantes brancos na pós-graduação.

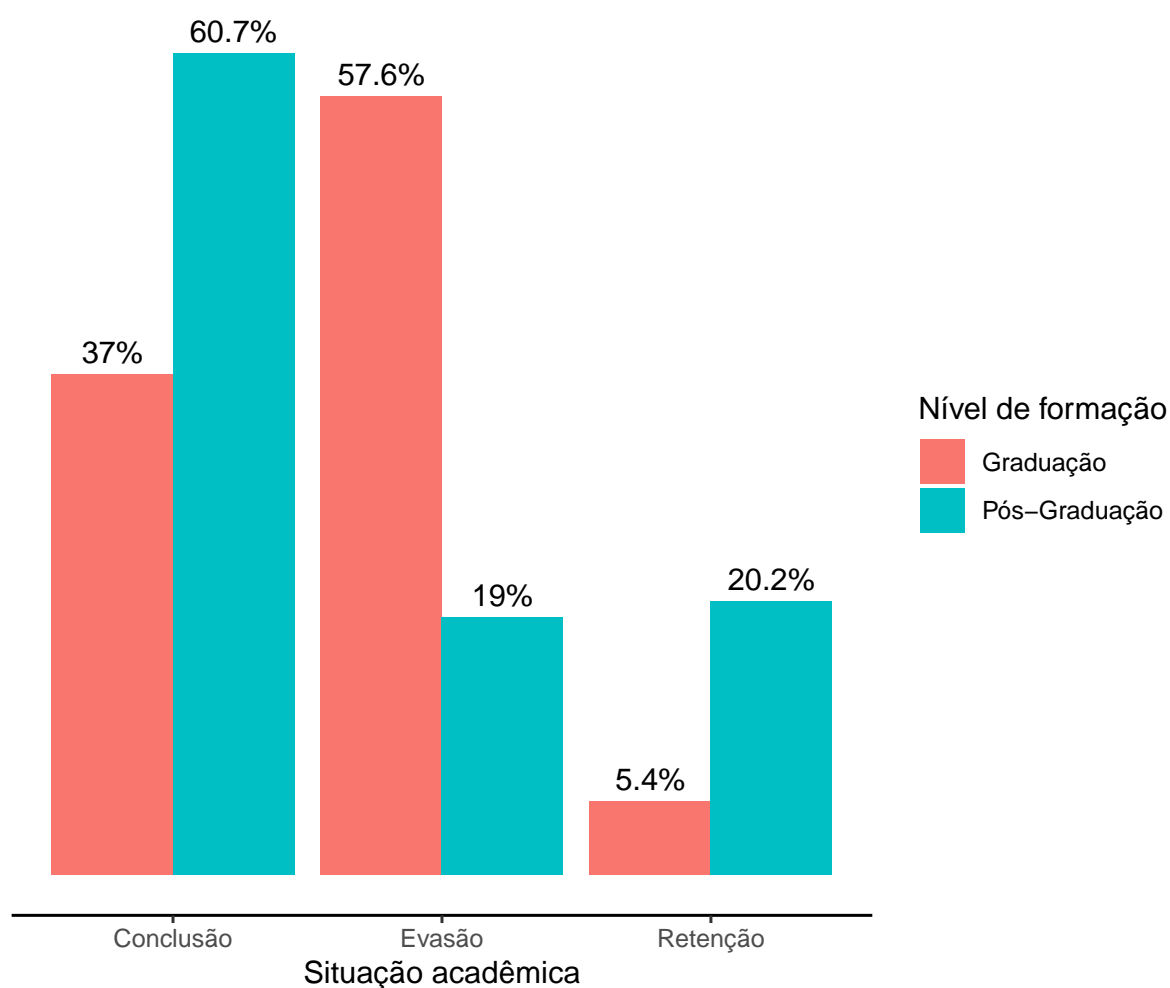


Figura 6 – Distribuição dos estudantes da UFV por situação acadêmica e nível de formação

Para a análise da trajetória acadêmica, foram consideradas três categorias fundamentais: conclusão, que representa o término do ciclo de estudos com a obtenção do grau; evasão, caracterizada pela interrupção do vínculo do estudante com o curso antes de sua finalização; e retenção, que se refere à permanência do aluno na instituição além do tempo regulamentar previsto para a formação.

Conforme ilustrado na Figura 6, nota-se que na graduação predomina a evasão (57,6%), seguida pela conclusão (37%) e pela retenção (5,4%). Já na pós-graduação, a maioria dos alunos conclui o curso (60,7%), enquanto 19% evadem e 20,2% permanecem retidos. Assim, observa-se que a pós-graduação apresenta uma taxa de conclusão significativamente maior e menor evasão em comparação à graduação.

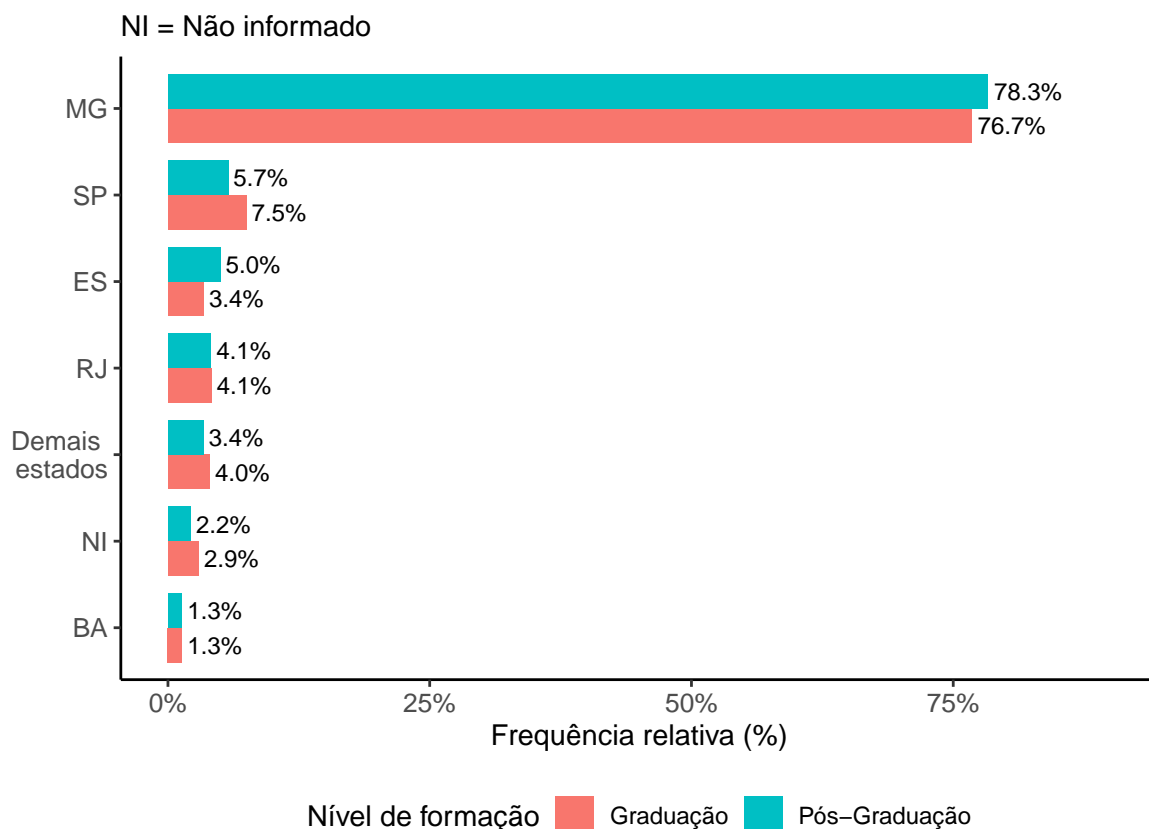


Figura 7 – Distribuição dos estudantes da UFV por UF e nível de formação

Em relação à distribuição dos estudantes segundo a unidade federativa de origem, o gráfico apresentado na Figura 7 indica uma expressiva concentração de discentes provenientes do estado de Minas Gerais, tanto na graduação quanto na pós-graduação, configurando-se como a principal origem dos alunos em ambos os níveis de formação. Nota-se que essa proporção é ligeiramente superior na pós-graduação, o que sugere maior permanência ou atração regional nesse nível acadêmico. As demais unidades federativas apresentam participações relativas bem inferiores, com destaque para São Paulo, Rio de Janeiro, Espírito Santo e Bahia, que figuram como os estados mais representativos fora de Minas Gerais. As categorias “Demais estados” e “NI” apresentam percentuais reduzidos, evidenciando a predominância de estudantes da área de influência da instituição e a boa completude da base de dados.

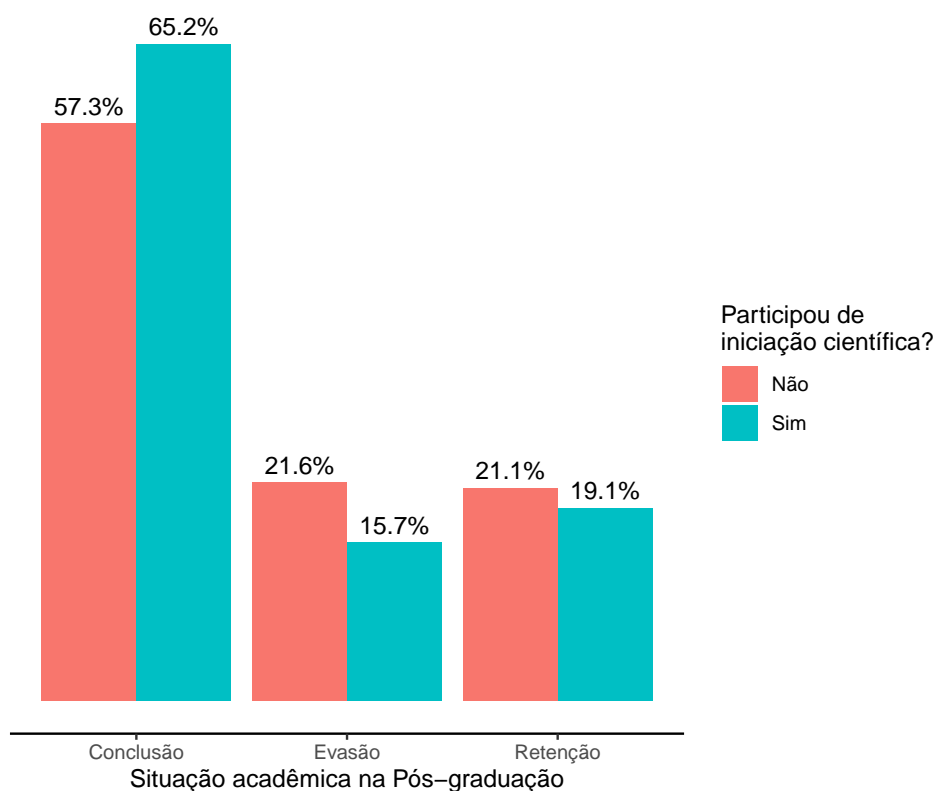


Figura 8 – Distribuição dos estudantes da UFV por situação acadêmica, nível de formação e IC

Na Figura 8, nota-se que, entre os alunos de pós-graduação, aqueles que participaram de Iniciação Científica (IC) apresentam maior taxa de conclusão (65,2%) em comparação com os que não participaram (57,3%). Por outro lado, a evasão é menor entre os participantes de IC (15,7%) do que entre os não participantes (21,6%), e a retenção apresenta pequena diferença entre os grupos (19,1% para IC versus 21,1% para não IC). Esses resultados indicam que a participação em IC está associada a melhores desfechos acadêmicos na pós-graduação.

Iniciação Científica	Ingressou na Pós-Graduação		Concluiu a Pós-Graduação	
	Não (44.306)	Sim (9.007)	Não (3.022)	Sim (5.985)
Não	40.108 (88%)	5.214 (12%)	1.973 (38%)	3.241 (62%)
Sim	4.198 (53%)	3.793 (47%)	1.049 (28%)	2.744 (72%)

Tabela 4 – Associação entre participação em iniciação científica, ingresso e conclusão da pós-graduação

A partir dos dados apresentados na Tabela 4, observa-se uma associação favorável entre a participação em programas de Iniciação Científica e o desempenho na pós-graduação. Estudantes que participaram de Iniciação Científica apresentam propor-

ções mais elevadas tanto de ingresso quanto de conclusão da pós-graduação em relação àqueles que não participaram, evidenciando o impacto positivo dessa experiência na continuidade e no êxito da trajetória acadêmica.

3.2 Regressão Logística — Ingresso na pós-graduação

Nesta seção, são apresentados os resultados do modelo de regressão logística ajustado para avaliar a associação entre a participação em Iniciação Científica e o ingresso na pós-graduação. A Tabela 5 apresenta a estimativa do coeficiente associado à Iniciação Científica, seu erro padrão, a razão de chances correspondente e o respectivo intervalo de confiança de 95%, bem como o nível de significância estatística. A adequação do modelo foi avaliada por meio do gráfico de envelope, utilizado como diagnóstico do ajuste.

Iniciação Científica	Coeficiente	Erro Padrão	OR	IC 95%	p-valor
Não (referência)	–	–	–	–	–
Sim	1,94	0,03	6,95	[6,59; 7,33]	< 0,01***

Coeficiente (β_1): estimativa do modelo de regressão logística; Erro Padrão: erro padrão do coeficiente estimado; OR: Razão de Chances; IC 95%: Intervalo de Confiança de 95%. *** significativo a 1%.

Tabela 5 – Estimativas do modelo de regressão logística para o ingresso na pós-graduação segundo participação em iniciação científica

Os resultados do modelo de regressão logística indicam que a participação em programas de Iniciação Científica está fortemente associada ao ingresso na pós-graduação, evidenciando um efeito estatisticamente significativo ao nível de 1%. Especificamente, estudantes que participaram de Iniciação Científica apresentaram uma razão de chances aproximadamente sete vezes maior de ingressar em cursos de pós-graduação quando comparados àqueles que não participaram, conforme indicado pelo valor da OR e pelo intervalo de confiança que não inclui a unidade. Esses números sugerem que a experiência em atividades de pesquisa durante a graduação desempenha papel relevante na continuidade da trajetória acadêmica, reforçando a importância institucional da Iniciação Científica como mecanismo de estímulo à formação científica e ao prosseguimento dos estudos em nível *stricto sensu*.

3.2.1 Avaliação do ajuste do modelo de ingresso na pós-Graduação

Na Figura 9, a análise do gráfico de envelope dos resíduos do modelo de regressão logística para o ingresso na pós-graduação indica um ajuste adequado aos dados observados. Observa-se que a grande maioria dos pontos se encontra dentro dos limites do envelope simulado, com apenas uma fração muito pequena das observações

posicionando-se fora desses limites (aproximadamente 0,01% do total), o que sugere ausência de desvios sistemáticos relevantes em relação às suposições do modelo. Além disso, a disposição dos resíduos ao longo dos quantis teóricos não evidencia padrões de assimetria ou caudas excessivamente pesadas, reforçando a adequação da especificação do modelo e a consistência das inferências realizadas a partir de suas estimativas.

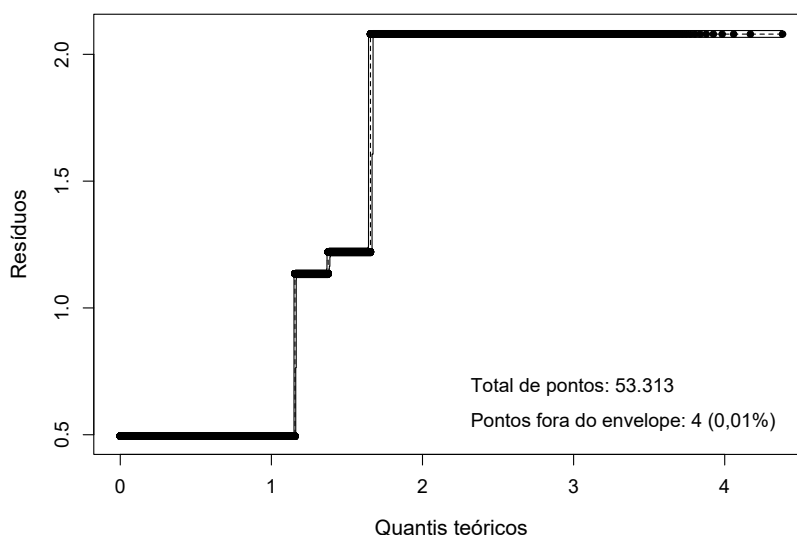


Figura 9 – Gráfico de envelope: ajuste dos resíduos do modelo de ingresso na pós-graduação

3.3 Regressão Logística — Conclusão da pós-graduação

Nesta seção, são apresentados os resultados do modelo de regressão logística ajustado para avaliar a associação entre a participação em Iniciação Científica e a conclusão da pós-graduação. A Tabela 6 apresenta a estimativa do coeficiente associado à Iniciação Científica, seu erro padrão, a razão de chances correspondente e o respectivo intervalo de confiança de 95%, bem como o nível de significância estatística. A adequação do modelo foi avaliada por meio do gráfico de envelope, utilizado como diagnóstico do ajuste.

Iniciação Científica	Coeficiente	Erro Padrão	OR	IC 95%	p-valor
Não (referência)	–	–	–	–	–
Sim	0,47	0,05	1,59	[1,45; 1,74]	< 0,01***

Coeficiente (β_1): estimativa do modelo de regressão logística; Erro Padrão: erro padrão do coeficiente estimado; OR: Razão de Chances; IC 95%: Intervalo de Confiança de 95%. *** significativo a 1%.

Tabela 6 – Estimativas do modelo de regressão logística para a conclusão da pós-graduação segundo participação em iniciação científica

Os resultados do modelo de regressão logística indicam que a participação em programas de Iniciação Científica está positivamente associada à conclusão da pós-graduação, apresentando efeito estatisticamente significativo ao nível de 1%. Estudantes que participaram de Iniciação Científica apresentam uma chance 59% maior de concluir a pós-graduação em comparação aos que não tiveram essa experiência prévia, conforme evidenciado pelo intervalo de confiança de 95% que se mantém integralmente acima da unidade. Esse resultado sugere que a experiência prévia em atividades de pesquisa contribui de forma relevante para a persistência e o êxito acadêmico na pós-graduação, ainda que com magnitude de efeito inferior à observada no modelo de ingresso, reforçando o papel da Iniciação Científica como elemento favorável ao desempenho acadêmico ao longo da trajetória educacional.

3.3.1 Avaliação do ajuste do modelo de conclusão da pós-graduação

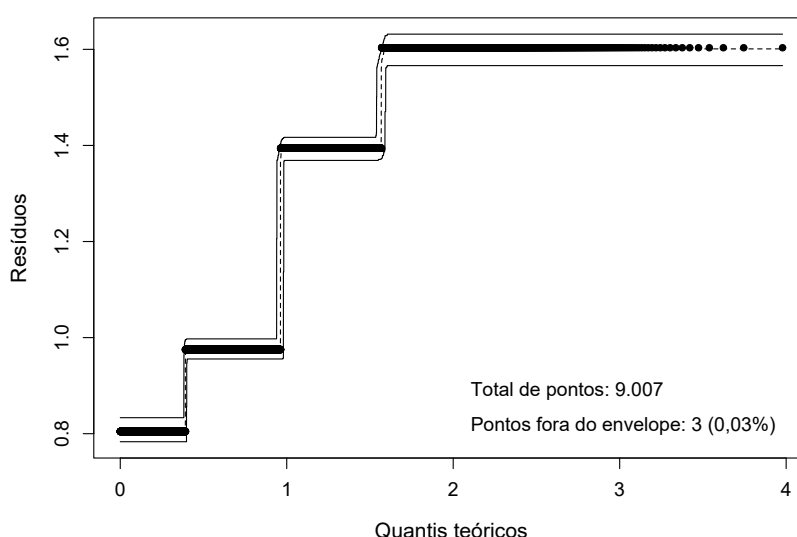


Figura 10 – Gráfico de envelope: ajuste dos resíduos do modelo de conclusão da pós-graduação

A análise do gráfico de envelope dos resíduos do modelo de regressão logística para a conclusão da pós-graduação indica que o ajuste do modelo aos dados é satisfatório. Observa-se que a quase totalidade das observações encontra-se contida dentro dos limites do envelope simulado, com apenas três pontos situando-se fora desses limites, o que corresponde a aproximadamente 0,03% do total de observações. Ademais, a disposição dos resíduos em relação aos quantis teóricos não evidencia padrões sistemáticos de desvio, assimetria acentuada ou comportamento anômalo nas caudas, sugerindo que as suposições do modelo são adequadamente atendidas e que as inferências realizadas a partir de suas estimativas podem ser consideradas consistentes e confiáveis.

3.4 Discussão

Os resultados deste estudo estão em consonância com a literatura nacional ao evidenciar uma associação positiva entre a participação em programas de Iniciação Científica e o ingresso, bem como a conclusão, de cursos de pós-graduação. Achados semelhantes são reportados por [Lordelo e Argôlo 2015](#), [Pelinski, Kampa et al. 2021](#), [Lima 2016](#), que destacam a Iniciação Científica como um fator relevante na preparação acadêmica e no fortalecimento do vínculo dos estudantes com a pesquisa científica. Assim, a maior probabilidade de continuidade e êxito acadêmico observada entre os estudantes com experiência prévia em pesquisa reforça o entendimento de que a Iniciação Científica contribui para o desenvolvimento de competências metodológicas e para uma inserção mais qualificada no ambiente acadêmico, alinhando-se ao padrão observado em diferentes instituições federais de ensino superior.

Nesse contexto, a análise da trajetória acadêmica não deve se limitar apenas ao ingresso, mas também à capacidade de conclusão dos cursos, uma vez que o fenômeno da evasão na pós-graduação brasileira tem sido objeto de preocupação crescente. Segundo [Fernandes et al. 2017](#), o mapeamento desses dados permite identificar gargalos institucionais, enquanto [Junior, Magalhães e Real 2020](#) argumentam que a gestão da evasão deve ser compreendida como uma política contínua, integrando métricas da graduação às da pós-graduação *stricto sensu*. A filtragem racial observada nos resultados da UFV corrobora a necessidade de políticas de permanência que considerem o recorte étnico-racial, de modo a possibilitar uma visão mais holística do fluxo discente e do impacto de experiências formativas prévias, como a Iniciação Científica, sobre a equidade na conclusão dos cursos.

3.5 Limitações do Estudo

A principal limitação desse estudo residiu no fato de que os dados são provenientes apenas da Universidade Federal de Viçosa. Nesse contexto, foram considerados apenas alunos que se graduaram na referida instituição ao se analisar o ingresso posterior na pós-graduação, bem como sua taxa de conclusão.

Assim, alunos que realizaram suas graduações em outras instituições e posteriormente decidiram prosseguir seus estudos em nível de pós-graduação na UFOP, não possuíam a informação de participação anterior em programas de pós-graduação. Esse fato pode ter gerado um efeito de subestimação das taxas de ingresso e conclusão da pós-graduação por alunos que possuíam a experiência prévia em iniciação científica.

4 Considerações Finais

Os resultados desta pesquisa permitem concluir que a Iniciação Científica exerce um papel estruturante na trajetória acadêmica dos estudantes, influenciando de forma significativa não apenas o ingresso, mas também a permanência e a conclusão de programas de pós-graduação. A experiência prévia em pesquisa mostrou-se associada a uma formação acadêmica mais consistente, favorecendo o desenvolvimento de habilidades metodológicas, do pensamento crítico e de uma maior familiaridade com a dinâmica da produção científica, aspectos fundamentais para o sucesso na pós-graduação. Ademais, a evidência empírica sugere que a Iniciação Científica atua como um mecanismo de seleção e preparação acadêmica, contribuindo para a redução das taxas de evasão e para a elevação dos índices de conclusão, o que reforça seu valor como política institucional de formação de recursos humanos qualificados.

Nesse contexto, os achados deste estudo subsidiam a formulação e o fortalecimento de políticas públicas e institucionais voltadas à ampliação do acesso à Iniciação Científica, especialmente em universidades públicas, além de destacarem a importância do uso sistemático de bases de dados institucionais para o monitoramento e a avaliação contínua de seus impactos. Por fim, os resultados contribuem para o avanço do debate acadêmico sobre a relação entre formação científica na graduação e desempenho na pós-graduação, oferecendo evidências robustas que podem orientar futuras pesquisas e ações estratégicas no âmbito do ensino superior brasileiro.

Estudos futuros podem inserir características sociodemográficas nos modelos de regressão logística, buscar alternativas para identificar se discentes graduados em outras instituições participaram de iniciação científica em suas instituições de origem, bem como a aplicação da metodologia proposta em outras instituições de ensino superior.

Referências

- AGRESTI, A. **Categorical Data Analysis**. 2. ed. New York: John Wiley & Sons, 2002. Citado na página 16.
- BASTOS, F. *et al.* A importância da iniciação científica para os alunos de graduação em biomedicina. **Revista Eletrônica Novo Enfoque**, v. 11, n. 11, p. 61–66, 2010. Citado na página 10.
- FERNANDES, E. F. *et al.* Panorama do fenômeno da evasão discente na pós-graduação: uma análise a partir do geocapes. In: **Anais do XVII Colóquio Internacional de Gestão Universitária**. Mar del Plata: [s.n.], 2017. p. 1–16. Citado na página 28.
- GORGENS, J. B. **Avaliação da produção científica dos egressos, bolsistas e não bolsistas de iniciação científica...** Dissertação (Mestrado) — Universidade Federal de Minas Gerais, 2007. Citado na página 10.
- JUNIOR, J. d. S. S.; MAGALHÃES, A. M. d. S.; REAL, G. C. M. A gestão da evasão nas políticas educacionais brasileiras: Da graduação à pós-graduação stricto sensu. **ETD Educação Temática Digital**, UNICAMP, v. 22, n. 2, p. 460–478, 2020. Citado na página 28.
- LIMA, L. G. A. d. **A influência da iniciação científica sobre a pós-graduação: um estudo de caso sobre tempo, idade de titulação e produção científica**. Dissertação (Mestrado) — Universidade Federal de São Carlos, 2016. Citado 2 vezes nas páginas 11 e 28.
- LORDELO, J. A. C.; ARGÔLO, R. F. de. Influências da iniciação científica na pós-graduação. **Estudos em Avaliação Educacional**, v. 26, n. 61, p. 168–191, 2015. Citado 2 vezes nas páginas 11 e 28.
- MASSI, L.; QUEIROZ, S. L. Estudos sobre iniciação científica no brasil: uma revisão. **Cadernos de Pesquisa**, Fundação Carlos Chagas, v. 40, n. 139, p. 173–197, 2010. Citado na página 10.
- MONTGOMERY, D. C. **Design and analysis of experiments**. [S.l.]: John Wiley & Sons, 2017. Citado na página 15.
- PELINSKI, J. D.; KAMPA, A. *et al.* Efeito da iniciação científica no rendimento acadêmico e na inserção em programas de pós-graduação “stricto sensu”. **Publicatio UEPG: Ciências Sociais Aplicadas**, 2021. Citado 2 vezes nas páginas 11 e 28.
- WARTON, D. I. Global simulation envelopes for diagnostic plots in regression models. **The American Statistician**, Taylor & Francis, v. 77, n. 4, p. 425–431, 2023. Citado na página 17.