

**UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
GRADUAÇÃO EM ESTATÍSTICA**

Gabriela Paschoal da Silva

**Produtividade Científica e Desigualdade de Gênero: Evidências nas Áreas de
Matemática e Estatística**

Juiz de Fora

2025

Gabriela Paschoal da Silva

**Produtividade Científica e Desigualdade de Gênero: Evidências nas Áreas de
Matemática e Estatística**

Trabalho de conclusão de curso apresentado ao Curso de Estatística da Universidade Federal de Juiz de Fora, como requisito parcial à obtenção do título de Bacharel em Estatística.

Orientadora: Prof^a. Dra. Camila Borelli Zeller

Coorientadora: Prof^a. Dra. Julia Castro Mendes

Juiz de Fora

2025

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

da Silva, Gabriela Paschoal .

Produtividade Científica e Desigualdade de Gênero: Evidências nas Áreas de Matemática e Estatística / Gabriela Paschoal da Silva. -- 2025.

83 f. : il.

Orientadora: Camila Borelli Zeller

Coorientadora: Julia Castro Mendes

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas, 2025.

1. Produtividade em pesquisa. 2. Desigualdade de gênero em STEM. 3. CNPq. 4. Testes de hipóteses não-paramétricos. 5. Regressão logística.. I. Zeller, Camila Borelli , orient. II. Mendes, Julia Castro , coorient. III. Título.

Gabriela Paschoal da Silva

**Produtividade Científica e Desigualdade de Gênero: Evidências nas Áreas de
Matemática e Estatística**

Trabalho de conclusão de curso apresentado ao Curso de Estatística da Universidade Federal de Juiz de Fora, como requisito parcial à obtenção do título de Bacharel em Estatística.

Aprovada em (dia) de (mês) de (ano)

BANCA EXAMINADORA

Prof^a. Dra. Camila Borelli Zeller - Orientadora
Universidade Federal de Juiz de Fora

Prof^a. Dra. Julia Castro Mendes - Coorientadora
Universidade Federal de Juiz de Fora

Prof^a. Dra. Bárbara da Costa Campos Dias
Universidade Federal de Juiz de Fora

Prof. Dr. Tiago Maia Magalhães
Universidade Federal de Juiz de Fora

AGRADECIMENTOS

Às condições universais aleatórias que resultaram na combinação das moléculas que me deram origem (afinal de contas, quais as probabilidades, não é mesmo?), juntamente a todas as outras combinações e acasos que me trouxeram até aqui.

Aos deuses, por terem regido todo o processo - e talvez este agradecimento contradiga o anterior, mas o ser humano é assim mesmo: cheio de contradições.

Às contradições. Afinal, ter optado pela Estatística como curso de segundo ciclo foi uma delas.

À minha família, pelo apoio e pelo amor. Em especial: à minha mãe Cristiana e ao meu pai Marcio. Sou grata pelo constante encorajamento aos meus estudos e por, desde cedo, me estimularem a buscar minha independência. Além disso, agradeço-lhes por terem me dado a liberdade para fazer minhas próprias escolhas, e por terem me apoiado nas grandes decisões que tomei em busca da realização dos meus sonhos.

À UFJF e ao governo federal, pelo ensino gratuito e de qualidade. Aos membros dos Departamentos de Ciências Exatas e de Estatística. Aos bons mestres, que fizeram a diferença.

Aos pesquisadores que me orientaram em projetos de Iniciação Científica - Dra. Camila Borelli, Dr. Clécio Ferreira, Dr. Glauco Carvalho, Dr. Alfredo Chaoubah - e ao CNPq, pelo financiamento (foram três projetos de IC ao total, incluindo um desenvolvido na Embrapa). Ao professor Dr. Jairo de Souza, pela orientação no estágio que realizei em sua equipe de Fluência em Leitura (CAEd/UFJF); e ao Dr. Tuko (Luiz Vicente Ribeiro), pela supervisão depois que fui transferida para a sua equipe de Instrumentos Contextuais e Indicadores.

Novamente à minha orientadora maravilhosa, Dra. Camila Borelli, que também foi minha primeira referência de mulher na Estatística. À minha coorientadora (também maravilhosa), Dra. Julia Mendes, juntamente à sua equipe de estudos em desigualdade de gênero do Programa de Pós-Graduação em Engenharia Civil (UFJF). Porque ao realizarem sua pesquisa com docentes da Engenharia, inspiraram o surgimento desta, que por sua vez se volta para a Estatística e para a Matemática. Além disso, agradeço o apoio que prestaram durante a realização deste trabalho.

Ao meu irmão Olavo, por ter me ajudado imensamente na coleta dos dados para este trabalho. Nesse aspecto, também agradeço as contribuições feitas por Ja-

queline e Maria Fernanda, companheiras de curso, e pelo professor Dr. Alexandre Cury (Engenharia Civil/UFJF). Agradeço também ao professor Dr. Marcel Vieira, por ter esclarecido minhas dúvidas acerca do procedimento adotado para a coleta dos dados neste estudo; e novamente ao professor Dr. Clécio Ferreira, por ter esclarecido as dúvidas referentes aos testes de hipóteses aplicados.

Por fim, agradeço à minha amiga Izabella pelo companheirismo e apoio que datam da época do vestibular. Caminhamos e crescemos juntas de lá para cá, e foi incrível. Em geral, agradeço aos amigos, colegas e a todos que conheci e com quem tive boas trocas nestes anos de faculdade, por terem feito de minha caminhada mais divertida, mesmo quando árdua.

“Se um dia te disserem que nossas histórias são pequenas, lembre-os que foi com mãos pequenas que modelei milhares de doces e versos.” (Cora Coralina em carta à filha, 1980).

“Ninguém nasce mulher: torna-se mulher.” (Simone de Beauvoir, 1949, p.3).

“A utopia está lá no horizonte. Me aproximo dois passos, ela se afasta dois passos. Caminho dez passos e o horizonte corre dez passos. Por mais que eu caminhe, jamais alcançarei. Para que serve a utopia? Serve para isso: para que eu não deixe de caminhar.” (Eduardo Galeano, 1993, p.230).

RESUMO

As bolsas de produtividade em pesquisa, concedidas pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) a pesquisadores(as) de destaque desde os anos 1970, são consideradas um dos principais programas de estímulo à ciência brasileira. Até 2024, essas bolsas eram divididas em três categorias (Sênior, categoria 1 e categoria 2) e a categoria 1 ainda era subdividida em quatro níveis distintos (A, B, C e D). Nesse sentido, este trabalho estudou a produção científica de pesquisadores(as) da área de Matemática e Estatística, com o objetivo de avaliar se é possível estabelecer critérios objetivos para classificá-los(as) como bolsistas - incluindo a especificação da categoria/nível - ou não-bolsistas. Para isso, utilizou-se a regressão logística sobre dados extraídos do *Lattes*, considerando variáveis como: quantidade de orientações realizadas, número de artigos publicados, participações em eventos, anos desde o doutoramento, etc. Além disso, considerando as barreiras sistêmicas enfrentadas por mulheres na carreira científica, buscou-se investigar as diferenças entre sexos na distribuição dos(as) pesquisadores(as), na produtividade acadêmica e na classificação feita pelo modelo. Nesse processo, foram aplicados testes não-paramétricos para caracterizar o perfil dos(as) pesquisadores(as) por produção, nível da bolsa (no caso dos(as) pesquisadores(as) bolsistas) e sexo, além de pré-selecionar variáveis para o modelo estatístico. A análise revelou que homens publicam mais artigos de prestígio enquanto mulheres se destacam em orientações de graduandos(as) e capítulos de livros - em geral, atividades menos reconhecidas. Além disso, evidenciou-se que as contribuições acadêmicas femininas são avaliadas com critérios mais rígidos, exigindo desempenho superior para alcançar o mesmo reconhecimento que seus pares masculinos. O estudo reforça a necessidade de políticas mais equitativas na ciência brasileira.

Palavras-chave: produtividade em pesquisa; desigualdade de gênero em stem; cnpq; testes de hipóteses não-paramétricos; regressão logística.

ABSTRACT

Since the 1970s, research productivity grants awarded by the National Council for Scientific and Technological Development (CNPq) have been considered one of the main programs for promoting science in Brazil. Until 2024, these grants were divided into three categories (Senior, Category 1 and Category 2), with Category 1 further subdivided into four distinct levels (A, B, C and D). In this context, the present study analyzed the scientific output of researchers in the field of Mathematics and Statistics, aiming to assess whether it is possible to establish objective criteria to classify them as grant recipients — including the specific category/level — or as non-recipients. To achieve this, logistic regression was applied to data extracted from the *Lattes* Platform, considering variables such as the number of student supervisions, published articles, participation in academic events, years since obtaining a Doctor of Philosophy degree (Ph.D.), among others. Moreover, taking into account the systemic barriers faced by women in scientific careers, the study also sought to investigate gender differences in the distribution of researchers, academic productivity and classification produced by the model. In this process, nonparametric tests were applied to characterize the profile of researchers according to productivity, grant level (for those receiving research grants), and gender, as well as to preselect variables for the statistical model. The analysis revealed that men publish more prestigious articles, while women stand out in mentoring undergraduate students and writing book chapters — activities that are generally less recognized. Furthermore, the findings show that women's academic contributions are evaluated more rigorously, requiring higher performance to achieve the same level of recognition as their male peers. The study reinforces the need for more equitable policies in Brazilian science.

Keywords: research productivity; gender inequality in stem; cnpq; non-parametric hypothesis tests; logistic regression.

LISTA DE ILUSTRAÇÕES

1	Quantitativo de docentes e bolsistas de produtividade em 2023.	19
2	Quantidade de artigos publicados por categoria QUALIS. Zero (0) representa não-PQs enquanto (1) representa PQs.	33
3	Total de artigos publicados (base toda, PQs e Não-PQs).	33
4	Publicações em congressos (base toda, PQs e Não-PQs).	34
5	Livros publicados (base toda, PQs e Não-PQs).	34
6	Capítulos de livros publicados (base toda, PQs e Não-PQs).	35
7	Orientações por categoria. Zero (0) representa não-PQs enquanto (1) representa PQs.	35
8	Anos desde doutoramento (PQs e Não-PQs).	36
9	Pós doutorados. Zero (0) representa não-PQs enquanto (1) representa PQs.	36
10	Participações em eventos (PQs e Não-PQs).	37
11	Distribuição de Pesquisadores(as) por Sexo e Status PQ.	40
12	Distribuição de Pesquisadores(as) PQ e Não-PQ por País de Origem. .	41
13	Distribuição de Não-PQs por Sexo para os Top 10 Países Estrangeiros.	42
14	Distribuição de PQs por Sexo para os Top 10 Países Estrangeiros. . .	43
15	Quantidade de artigos publicados por categoria QUALIS (Sexos Feminino e Masculino).	44
16	Total de artigos publicados (Base toda, Sexos Feminino e Masculino). .	44
17	Publicações em Congresso (Base Toda, Sexos Feminino e Masculino).	45
18	Número de livros publicados (Base Toda, Sexos Feminino e Masculino).	45
19	Número de capítulos de livros publicados (Base Toda, Sexos Feminino e Masculino).	46
20	Orientações por categoria (Sexos Feminino e Masculino).	46
21	Anos desde o doutoramento (Sexos Feminino e Masculino).	47
22	Quantidade de pós doutorados (Sexos Feminino e Masculino).	47
23	Participações em eventos (Sexos Feminino e Masculino).	48

LISTA DE TABELAS

1	CrITÉrios normativos para julgamento do projeto de pesquisa e distribuiçŁo das bolsas.	16
2	RelaçŁo das categorias e dos nÍveis das bolsas de produtividade em pesquisa (chamada N° 09/2023 do CNPq).	17
3	Indicadores estatisticamente (5%) diferentes entre PQs e NŁo-PQs. . .	38
4	DistribuiçŁo de Homens, Mulheres e Total na base de dados.	39
5	DistribuiçŁo de bolsas de produtividade por sexo.	41
6	Indicadores estatisticamente (5%) diferentes entre homens e mulheres (base toda).	49
7	Indicadores estatisticamente (5%) diferentes entre homens e mulheres (PQs).	51
8	DistribuiçŁo de bolsistas por nÍvel e gÊnero.	52
9	Indicadores estatisticamente diferentes (5%) entre homens e mulheres por categoria PQ.	53
10	EspecificaçŁo da regressŁo logÍstica binÁria ajustada.	56
11	Resultados da classificaçŁo feita com o modelo logÍstico binÁrio. . . .	58
12	VariÁveis explicativas do modelo logÍstico multinomial ajustado e p-valores associados (testes de <i>deviance</i> do tipo II).	61
13	Coeficientes e p-valores por indicador (variÁvel) e categoria de produtividade	62
14	Resultados da classificaçŁo feita com o modelo logÍstico multinomial. .	67
15	Métricas por classe para o modelo logÍstico multinomial.	67
A.1	Programas de Pós-GraduaçŁo contemplados	76
B.1	Indicadores estatisticamente (5%) iguais entre PQs e NŁo-PQs. . . .	79
B.2	Estatísticas para os indicadores de PQs e NŁo-PQs. Indicadores com asterisco (*) se mostraram significativos no teste da mediana.	79
C.1	Indicadores estatisticamente (5%) iguais entre homens e mulheres (caso geral).	81

C.2	Estatísticas para os indicadores de mulheres (F) e homens (H). Indicadores com asterisco (*) se mostraram significativos no teste da mediana.	81
C.3	Estatísticas para os indicadores significativos entre mulheres (F) e homens (H) bolsistas de produtividade (PQs).	83
C.4	Estatísticas para os indicadores significativos entre mulheres (F) e homens (H) por categoria da bolsa de produtividade.	83

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Sobre as Bolsas de Produtividade em Pesquisa	15
1.2	Sobre a Baixa Representatividade Feminina em STEM	18
2	OBJETIVOS	21
2.1	Objetivo Geral	21
2.2	Objetivos Específicos	21
3	FUNDAMENTAÇÃO TEÓRICA	22
3.1	Testes de Hipóteses	22
3.1.1	<i>Teste da Mediana</i>	22
3.1.2	<i>Teste Qui-Quadrado para Independência</i>	23
3.2	Regressão Logística	24
4	DESCRIÇÃO E PRÉ-PROCESSAMENTO DOS DADOS	28
4.1	Classificação do Sexo dos(as) Pesquisadores(as)	29
4.2	Classificação dos Pós-Doutorados	30
5	RESULTADOS DAS ANÁLISES DESCRITIVAS E INFERENCIAIS	31
5.1	PQs e Não-PQs	32
5.2	Status PQ e Sexo	39
5.2.1	<i>Diferenças entre Sexos (Juntando PQs e Não-PQs)</i>	43
5.2.2	<i>Diferenças entre Sexos (Apenas PQs)</i>	50
5.2.3	<i>Diferenças entre Sexos por Categoria e Nível PQ</i>	52
6	RESULTADOS DA MODELAGEM ESTATÍSTICA	54
6.1	Ajuste do Modelo Logístico Binário (PQs e Não-PQs)	54
6.1.1	<i>Classificação e Análise de Produtividade (PQs e Não-PQs)</i>	55
6.2	Ajuste do Modelo Logístico Multinomial (Categoria/Nível de PQ)	58
6.2.1	<i>Classificação e Análise de Produtividade (Categoria/Nível de PQ)</i>	60
7	DISCUSSÃO DOS RESULTADOS	68

7.1	Perfis Científicos e Desigualdade de Reconhecimento entre Sexos . . .	68
7.2	Sobre as Classificações (Modelagens)	69
7.3	Considerações Finais	70

1 INTRODUÇÃO

O Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), vinculado ao Ministério da Ciência, Tecnologia e Inovações (MCTI), é uma agência fundada em 1951 com o objetivo de promover e consolidar a pesquisa científica no Brasil (Silva, 2011). No início de sua criação, o CNPq contava com bolsas de estudo (ou formação) e de pesquisa, sendo posteriormente criadas as de iniciação científica, aperfeiçoamento ou especialização, estágio para desenvolvimento técnico, pesquisador(a) assistente, pesquisador(a) associado e chefe de pesquisa (Silva, 2011). O CNPq também conta com outros programas de alocação de recursos, como editais de fomento (ex. CNPq Universal), Programa de Apoio a Publicações Científicas, entre outros.

Em 2024, o CNPq aplicou um total de R\$ 2,6 bilhões em fomento, e em maio de 2025 contava com 91 mil bolsas vigentes, 15 mil projetos em curso e 8 milhões de currículos cadastrados na plataforma, de acordo com o seu site oficial (CNPq, 2025).

1.1 Sobre as Bolsas de Produtividade em Pesquisa

O programa de bolsas de Produtividade em Pesquisa (PQ) foi instituído pelo CNPq em 1976 com o objetivo de incentivar a produção de cientistas brasileiros(as), de forma a contribuir para a consolidação e maturação da Ciência no país (Silva, 2011). Segundo o site oficial do CNPq, a finalidade do programa é distinguir pesquisadores(as), enaltecendo sua produção científica (CNPq, 2025). Para concorrer ao auxílio, é necessário que o(a) pesquisador(a) interessado(a) possua título de doutor, currículo ativo na Plataforma Lattes, projeto de pesquisa e vínculo formal com a instituição onde será realizado o projeto. As bolsas PQ são distribuídas segundo critérios normativos estabelecidos pelo próprio CNPq, além de critérios específicos definidos pelos Comitês de Assessoramento de cada área de pesquisa.

Considerando os critérios de julgamento, a nota final dada a cada proposta é a média ponderada das notas atribuídas a cada item. Considerando a Tabela 1, percebe-se que o item de maior peso é o currículo do(a) pesquisador(a) (item 1). Depois, vem a relevância e originalidade do projeto de pesquisa (item 2), seguido pela orientação de alunos(as) (item 3) e, por fim, questões de internacionalização e *networking* (itens 4 e 5). Apesar disso, na prática, a distribuição das bolsas se apoia fortemente na produti-

Tabela 1: Critérios normativos para julgamento do projeto de pesquisa e distribuição das bolsas.

Critérios para Julgamento	Peso	Nota
1. Relevância, originalidade e caráter inovador da contribuição científica, tecnológica, intelectual e artística do proponente ao longo da carreira, com ênfase na atividade recente (últimos 5 anos).	3	0 a 10
2. Mérito acadêmico e intelectual, originalidade e relevância do projeto para o desenvolvimento científico, tecnológico ou social do país, considerando, adicionalmente, seus potenciais impactos e caráter inovador.	2	0 a 10
3. Contribuição do proponente para a formação de recursos humanos altamente qualificados e atuação em cursos de graduação e pós-graduação.	2	0 a 10
4. Cooperação com grupos de pesquisas ou instituições no país e no exterior e participação ou coordenação de projetos e redes de pesquisa.	2	0 a 10
5. Atuação em sociedades científicas e editoria de periódicos no país e no exterior, atuação em gestão científica, prêmios, condecorações e outras distinções, considerando sua fase profissional.	1	0 a 10

Fonte: Chamada CNPq N° 09/2023.

vidade do(a) pesquisador(a), sendo pouco provável que um desempenho considerado insuficiente receba investimento (Ibañez, 1996).

A rigor, as bolsas de Produtividade em Pesquisa (PQ) e Produtividade em Pesquisa Sênior (PQ-Sr) são duas linhas diferentes do mesmo programa, sendo a primeira (bolsa PQ) destinada a pesquisadores(as) que se destacam entre seus pares, respeitando os critérios normativos (Tabela 1) e critérios específicos definidos pelos Comitês de Assessoramento; e a segunda (bolsa PQ-Sr), destinada a pesquisadores(as) que se destacam entre seus pares como líderes e paradigmas em suas áreas de atuação, também respeitando os critérios normativos e critérios definidos pelos Comitês de Assessoramento (CNPq, chamada n° 09/2023). Apesar disso, neste trabalho, as duas linhas serão tratadas simplesmente como bolsas PQ de diferentes categorias. Assim, até a chamada n° 09/2023, as bolsas PQ eram classificadas em três categorias (Sê-

nior, 1 e 2) e a categoria 1 era subdividida em quatro níveis distintos (A, B, C e D). A relação das categorias e dos níveis é apresentada na Tabela 2.

Tabela 2: Relação das categorias e dos níveis das bolsas de produtividade em pesquisa (chamada Nº 09/2023 do CNPq).

Categoria	Sênior	1	2
Nível	—	A, B, C, D	—

Fonte: Elaborada pela autora.

Dentre os critérios de elegibilidade para as bolsas nas categorias 1 e 2 no edital de 2023, constam: ter obtido o título de doutor(a) até o ano de 2015 (categoria 1) e ter obtido o título de doutor(a) até o ano de 2020 (categoria 2). Além disso, para ambas as modalidades, o(a) pesquisador(a) não devia possuir bolsa PQ ou DT (Produtividade em Desenvolvimento Tecnológico e Extensão Inovadora) ativas, com vigência ultrapassando o ano de 2024.

Já no caso da bolsa PQ na categoria Sênior, os seguintes critérios deveriam ser observados para a elegibilidade: ter sido bolsista PQ ou DT na categoria 1 por pelo menos 20 anos, consecutivos ou não, ou ter sido bolsista PQ ou DT na categoria 1, níveis A ou B, por pelo menos 15 anos, consecutivos ou não. Além disso, o(a) pesquisador(a) deveria estar ativo no desenvolvimento de ciência/tecnologia e na formação de pesquisadores(as) em diversos níveis.

Em relação às categorias 1 e 2 (Tabela 2), para cada proposta recomendada para a aprovação, o Comitê Assessor é responsável por sugerir a categoria e o nível da bolsa de produtividade concedida para o(a) proponente pelo CNPq. Já em relação à categoria Sênior, foi feita apenas a recomendação da aprovação ou não da proposta. Essas sugestões são então revisadas pela área técnico-científica do CNPq, e a decisão final é dada pelo(a) presidente(a) do Conselho. No resultado público, constaram as propostas aprovadas, que por sua vez foram contratadas considerando o limite orçamentário (CNPq, chamada nº 09/2023).

Na chamada nº 09/2023, a vigência das bolsas se iniciou a partir do dia primeiro de março de 2024. A duração concedida para as bolsas PQ na categoria Sênior e na categoria/nível 1A foi de até 60 meses (5 anos); nas categorias/níveis 1B, 1C e 1D, de até 48 meses (4 anos); e na categoria 2, de até 36 meses (3 anos).

1.2 Sobre a Baixa Representatividade Feminina em STEM

Até 2030, estima-se que 20 milhões de empregos serão automatizados no Brasil, mas as profissões das áreas de Ciência, Tecnologia, Engenharia e Matemática (STEM, em inglês) estão entre as que mais crescerão - os dados são do McKinsey Global Institute. O problema disso é que as mulheres estão sub-representadas nessas áreas: em 2023, apesar de representarem 52% dos(as) candidatos(as) no Exame Nacional do Ensino Médio (Enem), apenas 12% das mulheres escolheram áreas da tecnologia e 18% optaram pela engenharia. Em contrapartida, 70% das mulheres escolheram cursos da Saúde e das Ciências Sociais (McKinsey & Company, 2025).

Ao contrário do que é estabelecido no senso comum, a tímida presença feminina em STEM não ocorre devido a falta de interesse ou de capacidade inatas, mas sim por questões culturais e históricas estruturais. De acordo com a psicóloga Carol Dweck (Uol, 2025), as crenças de um indivíduo acerca de suas habilidades começam na infância. Nesse período, existe uma influência muito grande de estereótipos de gênero que dão a entender, direta ou indiretamente, que áreas como engenharia, ciência e tecnologia são masculinas (Uol, 2025). Esse entendimento é alimentado pela falta de representatividade feminina nessas áreas, o que contribui para um ciclo de afastamento das mulheres das exatas (Lopes et al., 2024).

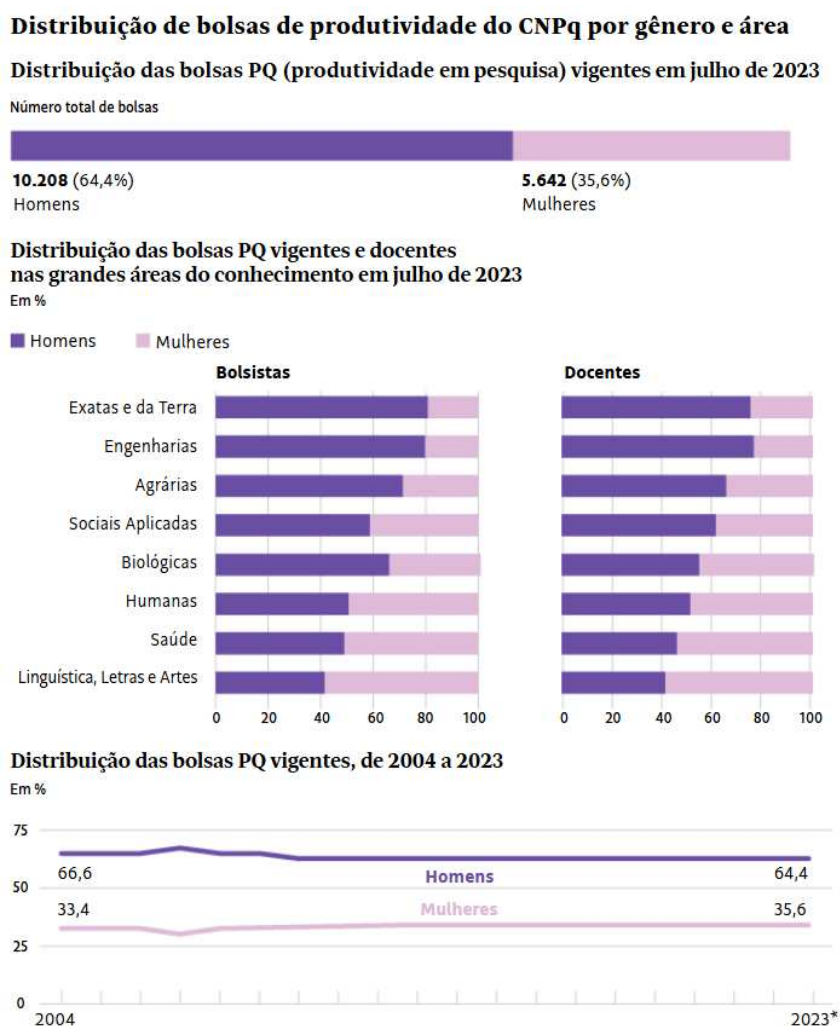
Em 2023, a plataforma de impacto social Força Meninas conduziu um estudo com cerca de 1400 garotas de várias regiões do Brasil. Intitulada "Meninas Curiosas, Mulheres de Futuro", a pesquisa revelou que 60% das entrevistadas disseram se sentir desmotivadas a seguir carreiras em STEM; 45% relataram acreditar que essas profissões são predominantemente masculinas e 62% disseram não conhecer profissionais nessas áreas (Força Meninas, 2023).

As tendências observadas no estudo realizado pela Força Meninas são refletidas nas instituições de ensino superior (IES) de todo o país: em Engenharias e em Ciências Exatas e da Terra, a presença masculina beira os 80%, de acordo com dados do CNPq referentes a 2023. Já em Saúde e em Linguística, Letras e Artes, o percentual masculino cai para cerca de 40%.

A desigualdade de gênero existente no corpo docente das instituições de ensino brasileiras, por sua vez, repercute na distribuição das bolsas de produtividade. De acordo com o CNPq, em julho de 2023, menos de 15% dos(as) docentes de IES nacionais

eram bolsistas de produtividade. Desses(as) bolsistas, 36% eram mulheres e 64% eram homens (Botallo, 2024). A relação do quantitativo de docentes e bolsistas PQ, por gênero e área do conhecimento, pode ser verificada na Figura 1.

Figura 1: Quantitativo de docentes e bolsistas de produtividade em 2023.



Fonte: Folha de S. Paulo (2024).

A questão da sub-representatividade feminina em STEM é ainda mais acentuada para as mulheres negras, apesar de mais da metade da população brasileira ser autodeclarada preta ou parda (IBGE, 2022). Em um artigo publicado pela Comissão de Gênero e Diversidade, da Sociedade Brasileira de Matemática (SBM) e da Sociedade Brasileira de Matemática Aplicada e Computacional (SBMAC), em parceria com a Associação Brasileira de Estatística (ABE), teve-se por objetivo identificar o perfil de estudantes do ensino superior em Matemática, Matemática Aplicada e Estatística em

relação ao sexo e à raça. Nisso, foi constatado que as mulheres negras apresentaram a menor participação quando comparadas à proporção total de mulheres (Queiroz, 2023). O estudo foi feito com dados de 2009 a 2019 extraídos do Censo da Educação Superior (Censup), vinculado ao Ministério da Educação (MEC). Além disso, de acordo com uma pesquisa do Grupo de Estudos Multidisciplinares da Ação Afirmativa da Universidade do Estado do Rio de Janeiro (Gemaa/UERJ), no ano de 2023, as mulheres negras e indígenas eram apenas 2,5% dos(as) professores(as) em programas de pós-graduação nas Ciências Exatas, da Terra e Biológicas (Freire, 2025). Já as mulheres brancas representavam 29% do total de professores(as) e os homens brancos, 61%, sendo este último grupo o majoritário.

2 OBJETIVOS

Analisando dados de profissionais bolsistas e não bolsistas de Produtividade em Pesquisa do Comitê de Assessoramento de Matemática e Estatística, o estudo buscou identificar critérios objetivos de distinção entre os dois grupos, no intuito de verificar se é possível classificar pesquisadores(as) em PQs ou não-PQs a partir desses critérios. Além disso, considerando as categorias de bolsa de produtividade existentes, buscou-se também identificar aspectos que distinguem pesquisadores(as) bolsistas de diferentes categorias. Ademais, considerando as desigualdades históricas de gênero em STEM, este estudo analisou a distribuição de homens e mulheres na base de dados e, adicionalmente, investigou se existem diferenças de produtividade e de classificação entre os sexos feminino e masculino, com foco no grupo de pesquisadores(as) PQs.

2.1 Objetivo Geral

O objetivo geral consiste em encontrar critérios objetivos que distingam pesquisadores(as) não-bolsistas de pesquisadores(as) bolsistas de produtividade - no segundo caso, especificando a categoria da bolsa (1A, 1B, 1C, 1D, 2). A categoria Sr (sênior) não será considerada por se tratar de um grupo de pesquisadores(as) com muitos anos de experiência, podendo apresentar critérios de produtividade mais subjetivos. Serão considerados(as) apenas pesquisadores(as) da área de Matemática, Probabilidade e Estatística no Brasil.

2.2 Objetivos Específicos

Este estudo possui como objetivos específicos: (1) analisar, em termos de gênero, o conjunto de dados utilizado; (2) investigar, em relação às variáveis analisadas, as diferenças entre pesquisadores (homens) e pesquisadoras (mulheres); (3) analisar a especificação dos modelos ajustados, bem como os resultados da classificação realizada na base de teste. Esse último objetivo tem por intuito verificar se homens são superestimados em relação às mulheres, além de buscar identificar vieses de gênero possivelmente refletidos pelos coeficientes e interações do modelo.

3 FUNDAMENTAÇÃO TEÓRICA

Para a realização deste estudo, foram aplicadas técnicas estatísticas e computacionais variadas, todas amplamente encontradas na literatura. A fundamentação teórica por trás das principais técnicas adotadas na metodologia será abordada de forma breve nesta seção, enquanto as demais serão devidamente referenciadas no trabalho.

3.1 Testes de Hipóteses

Neste trabalho, precisaremos verificar se há diferenças nas tendências centrais de diferentes grupos (PQs e Não-PQs, homens e mulheres). Por isso, será aplicado o teste não-paramétrico da mediana. A escolha de testes não-paramétricos se justifica pela natureza de contagem da maior parte das variáveis consideradas no estudo. Além disso, para verificar se há evidências de associação entre sexo e recebimento de bolsa de produtividade, será aplicado o teste qui-quadrado para independência. Para todos os testes aplicados, o nível de significância considerado foi o de 5%.

3.1.1 *Teste da Mediana*

O teste da mediana foi criado para comparar duas amostras independentes X e Y e verificar se elas vêm da mesma distribuição em relação à mediana. Mais especificamente, utiliza-se a mediana combinada como ponto de corte, e então realiza-se um teste (qui-quadrado ou Fisher) em uma tabela de contingência 2x2. Esta é uma adaptação de um teste aproximado para várias amostras independentes (Conover, 1999), já que um teste exato é de difícil alcance (Gibbons e Chakraborti, 2003).

Para aplicação do teste, foi utilizada uma função que se baseia na matriz esquematizada a seguir (essa matriz é a representação computacional da tabela de contingência 2x2):

$$\begin{bmatrix} \text{Acima (X)} & \text{Abaixo/igual (X)} \\ \text{Acima (Y)} & \text{Abaixo/igual (Y)} \end{bmatrix},$$

onde “Acima (X)” representa a quantidade de observações da amostra X que ficaram

acima da mediana combinada; “Abaixo/igual (X)” representa a quantidade de observações de X que ficaram abaixo da (ou igual à) mediana combinada. Similarmente, faz-se as interpretações para Y. O algoritmo utilizado (de autoria do professor Clécio Ferreira - Estatística/UFJF) é apresentado no Anexo D.

Com base na matriz descrita acima, se o número total de observações for menor que 20, o Teste Exato de Fisher é aplicado. Caso contrário, utiliza-se o Teste Qui-Quadrado para Independência, assumindo que as frequências são suficientemente grandes. As hipóteses do teste da mediana são definidas como

$$H_0: \theta_x = \theta_y$$

$$H_1: \theta_x \neq \theta_y,$$

onde θ_x é a mediana do primeiro grupo (X) e θ_y é a mediana do segundo grupo (Y).

3.1.2 Teste Qui-Quadrado para Independência

Considerando uma tabela de contingência 2x2 para variáveis categóricas - digamos, X e Y -, podemos aplicar um teste para saber se há evidências de associação entre as duas variáveis: o teste qui-quadrado para independência. Seja p_{ij} a probabilidade conjunta para as duas variáveis e p_{i+} e p_{+j} , as probabilidades individuais (para X e Y, respectivamente). Então, as hipóteses do teste são definidas como

$$H_0: p_{ij} = p_{i+}p_{+j}$$

$$H_1: p_{ij} \neq p_{i+}p_{+j},$$

onde a hipótese nula (H_0) é de independência estatística entre X e Y (Agresti, 2007). Para que esse teste seja aplicável, as frequências esperadas da tabela devem ser maiores que 5 (Everitt, 1992). O teste qui-quadrado para independência é baseado na estatística de mesmo nome, dada por

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - E_{ij})^2}{E_{ij}},$$

onde r é o número de linhas da tabela de contingência, c é o número de colunas, n_{ij} é a frequência observada na célula da linha i e coluna j , e E_{ij} é a frequência esperada na célula (i,j) , assumindo independência de X e Y (Agresti, 2007).

Como fixamos $r = c = 2$, o número de graus de liberdade associado à estatística é igual a $(r-1)(c-1) = (2-1)(2-1) = 1$.

3.2 Regressão Logística

Neste estudo, existe o interesse de verificar se é possível distinguir os(as) pesquisadores(as) PQs dos Não-PQs utilizando as variáveis de produtividade. Para isso, será adotado o modelo de regressão logística, que se baseia na premissa de que a variável resposta é binária (indicando a presença ou ausência do atributo de interesse) e de que as observações são independentes (Hosmer et al, 2013). Seja Y uma variável resposta e x , um valor observado para uma variável explicativa. Então a quantidade $E(Y|x)$ é o valor esperado de Y dado x . Na regressão linear, essa média é expressa como uma equação linear em x , o que implica que Y pode assumir qualquer valor conforme x varia no intervalo $(-\infty, +\infty)$. Essa equação pode ser expressa por

$$E(Y | x) = \beta_0 + \beta_1 x,$$

onde β_0 é o intercepto da reta (valor esperado de Y quando $x=0$), β_1 é o coeficiente angular (o quanto Y varia, em média, para cada unidade de x) e x é a variável explicativa (Montgomery, 2012).

Quando a regressão utilizada é a logística (ou seja, quando Y é dicotômico), $E(Y|x) = \pi(x)$ e a forma do modelo é

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}},$$

com $\pi(x) \in [0, 1]$, sendo $\pi(x) = P(Y = 1|x)$ a probabilidade de ocorrência da classe “positiva” da variável resposta dicotômica. Nesse modelo, pode ser aplicada a transformação logística, isto é

$$g(x) = \ln\left[\frac{\pi(x)}{1 - \pi(x)}\right] = \beta_0 + \beta_1 x,$$

onde $g(x)$ é chamado de logito e é linear em seus parâmetros, podendo ser contínuo, dependendo do intervalo no qual x está definido (Hosmer et al, 2013).

Além disso, o modelo de regressão logística apresenta outros pressupostos, tais como: a relação entre as variáveis explicativas contínuas e o logito da variável resposta deve ser linear; os erros devem apresentar distribuição binomial com média 0 e variância $\pi(x)[1-\pi(x)]$; as variáveis explicativas não devem ser altamente correlacionadas entre si (ausência de multicolinearidade) (Hosmer et al, 2013).

As equações apresentadas até aqui são referentes à regressão logística simples, que utiliza apenas uma variável explicativa. A generalização para mais de uma variável é a regressão logística múltipla, cuja fórmula é denotada por

$$g(\mathbf{x}) = \ln\left[\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p,$$

onde:

- $\pi(\mathbf{x}) = P(Y = 1|\mathbf{x})$ é a probabilidade de ocorrência do evento de interesse;
- $\mathbf{x} = (x_1, x_2, \dots, x_p)^T$ é o vetor de variáveis explicativas;
- $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$ é o vetor de coeficientes;
- β_0 é o intercepto (constante).

O intercepto β_0 representa o logaritmo da chance (*odds*) de ocorrência da classe positiva quando todas as variáveis explicativas assumem valor zero. Já os demais coeficientes representam o aumento (ou a redução) multiplicativo(a) nas chances de ocorrência da classe positiva, para cada unidade acrescida às variáveis explicativas correspondentes, mantendo as demais constantes. Por exemplo, se $\beta_1 > 0$, então, para cada unidade acrescida a x_1 , a chance de ocorrência da classe positiva é multiplicada por $\exp(\beta_1)$, mantendo-se as demais variáveis explicativas constantes e considerando um contexto sem interações entre variáveis.

Além dos efeitos principais, o modelo de regressão logística pode incorporar termos de interação entre variáveis explicativas, o que permite investigar se o efeito de uma variável sobre a chance do evento de interesse depende do valor de outra. A forma geral do modelo com duas variáveis explicativas, além da interação entre elas, é

$$\ln\left[\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 x_2).$$

Neste caso,

- β_1 representa o efeito de x_1 quando $x_2 = 0$;
- β_2 representa o efeito de x_2 quando $x_1 = 0$;
- β_3 representa a modificação do efeito de x_1 em função de x_2 (ou vice-versa) - mais precisamente, ele indica o quanto o efeito de x_1 nas chances do evento muda para cada unidade adicional de x_2 .

Assim, a presença de um termo de interação implica que o efeito de uma variável depende da outra, ou seja, os efeitos não são mais constantes, mas variam conjuntamente.

Neste trabalho, estaremos interessadas em estudar o efeito de indicadores de produtividade em conjunto com o sexo dos(as) pesquisadores(as), sendo atribuído o valor 1 ao sexo feminino e 0 ao masculino. Assim, supondo que x_1 represente a variável “sexo” e x_2 , a variável “número de artigos escritos”, teremos as seguintes interpretações: para os homens, a chance de ser PQ sofre um aumento multiplicativo de $\exp(\beta_2)$ para cada artigo adicional. Já para as mulheres, esse aumento é de $\exp(\beta_2 + \beta_3)$ por artigo, ou seja, o efeito de x_2 varia conforme o sexo do(a) pesquisador(a).

Por fim, considerando apenas pesquisadores(as) PQ, neste estudo também estaremos interessadas em modelar as categorias de produtividade. Para isso, será adotado o modelo logístico multinomial, que funciona da mesma maneira que o modelo logístico binário, já que estabelece comparações entre pares de categorias da variável resposta (Field et al, 2012). Nesse modelo, uma das categorias da variável dependente é escolhida como a referência, e as comparações são estabelecidas entre essa referência e as demais categorias (Field et al, 2012). Neste trabalho, a categoria

2 de produtividade será fixada como a referência e será comparada com as demais categorias (1D vs 2, 1C vs 2, 1B vs 2 e 1A vs 2).

4 DESCRIÇÃO E PRÉ-PROCESSAMENTO DOS DADOS

Como etapa inicial, buscou-se fazer o *download* manual de currículos *Lattes* (no formato XML) de docentes permanentes de programas de pós-graduação (PPGs) da área “Matemática, Probabilidade e Estatística”. Essa área foi escolhida por afinidade, pois é o campo de atuação da autora e da orientadora deste trabalho. A relação dos PPGs contemplados por IES e nível do programa (CAPES, 2021) pode ser conferida na Tabela A.1, contida no Apêndice A. Os PPGs da Fundação Universidade Federal de Sergipe (FUFSE) e da Universidade Federal do Paraná (UFPR) foram desconsiderados devido a problemas em seus respectivos *websites*, o que impossibilitou o acesso à relação de seus (suas) docentes. Além disso, o *website* do PPG em Matemática da UFMG não fazia distinção entre docentes permanentes e colaboradores(as) - desse modo, pode ser que tenham sido coletados currículos de pesquisadores(as) de ambos os grupos. Assim, dos 1410 nomes vinculados aos PPGs de “Matemática, Probabilidade e Estatística” (encontrados por buscas nos sites dos PPGs), foram localizados 1404 currículos na plataforma *Lattes*.

Para manter a consistência na análise dos(as) bolsistas PQ, não foram considerados(as) pesquisadores(as) bolsistas dos programas Produtividade em Desenvolvimento Tecnológico e Extensão Inovadora (CNPq), Desenvolvimento Tecnológico Industrial (CNPq) e Produtividade FAPEMIG/CNPq - consequentemente, 5 pessoas foram excluídas do estudo. Ademais, outros(as) 16 pesquisadores(as), por serem bolsistas de produtividade da categoria Sênior (categoria com critérios de avaliação diferentes), foram omitidos(as) do estudo. Dessa forma, trabalhamos com os dados de 1383 pesquisadores(as) ao total, sendo 427 PQs e 956 Não-PQs. O procedimento de *download* dos currículos começou no dia 18 de outubro de 2024 e levou cerca de três semanas.

Após a coleta de currículos, foi utilizado um código desenvolvido pelo professor Alexandre Abrahão Cury (Engenharia Civil - UFJF) no *software* Matlab® para extrair do XML de cada pesquisador(a) os seguintes dados (associados aos seguintes atributos de interesse):

- O número de artigos com QUALIS (CAPES, 2023) publicados em periódicos A1, A2, A3, A4, B1, B2, B3, B4 e C, respectivamente, além do total de artigos com

QUALIS publicados (produção científica);

- Número de livros e de capítulos de livros publicados (produção científica);
- Número de orientações para: Iniciação Científica (IC) em andamento e concluída; Trabalho de Conclusão de Curso (TCC); mestrado em andamento e concluído; doutorado em andamento e concluído (formação de recursos humanos);
- Participação em eventos científicos (engajamento profissional);
- Número de pós-doutorados feitos no Brasil e no exterior, além do total de pós-doutorados (qualificação suplementar e internacionalização);
- Ano do doutoramento (tempo de experiência profissional).
- Nacionalidade;

Em respeito a questões éticas, o algoritmo desenvolvido no *Matlab* codifica os nomes dos(as) pesquisadores(as), de forma que estes passaram a ser identificados(as) por códigos alfanuméricos (PQ-1A-1, PQ-1A-2, PQ-2-1, PQ-2-2, PQ-2-3, etc). A classificação QUALIS dos artigos segue o quadriênio 2017-2020, o mais recente disponível no momento da coleta dos dados. Artigos publicados em revistas sem QUALIS foram classificados dessa maneira e deixados de fora do presente estudo, independentemente do fator de impacto. Essa decisão se justifica pela dificuldade de fazer equivalências entre a classificação QUALIS e classificações baseadas em outros critérios. Ademais, o *software* R® de programação foi utilizado como ferramenta para a realização de todas as análises descritivas, inferenciais e preditivas conduzidas neste estudo.

4.1 Classificação do Sexo dos(as) Pesquisadores(as)

Neste estudo, optou-se pelo uso do termo “sexo” para designar a atribuição do sexo biológico no nascimento em detrimento do termo “gênero”, empregado quando a referência é à identidade social das pessoas. Até o momento de realização do *download* de currículos, a informação sobre o sexo dos(as) pesquisadores(as) não era disponibilizada pela plataforma *Lattes*. Portanto, foi empregado um procedimento de classificação do sexo com base nos nomes dos(as) docentes, abordagem essa que já foi

adotada em pesquisas semelhantes (Nery, 2025). Dessa forma, o primeiro nome de cada docente foi extraído e fornecido ao aplicativo GenderAPI® e ao Chat GPT®, ferramentas utilizadas para a classificação. Nessa etapa, obteve-se uma concordância de 98% entre as ferramentas. Para os indivíduos associados às discordâncias, foram pesquisados os currículos e a atribuição do sexo foi feita com base no pronome utilizado pelo(a) pesquisador(a). Esse processo, juntamente à divisão binária adotada (masculino/feminino), constitui uma limitação deste trabalho, já que não leva em consideração a identidade de gênero de cada indivíduo, se baseando apenas nos nomes (ou nos pronomes) para efetuar a classificação. No entanto, em estudos com grandes conjuntos de dados sem rótulo próprio, a informação mais confiável disponível para inferência do gênero é, em muitas das vezes, o nome (Santamaría e Mihaljević, 2018).

4.2 Classificação dos Pós-Doutorados

A inteligência artificial (IA) Grok® foi utilizada para classificar os pós-doutorados em nacionais e internacionais. Para isso, foi fornecido a ela um arquivo XLSX contendo uma coluna de códigos (associados aos (às) pesquisadores(as)) e outra coluna contendo o nome da instituição na qual foi feito o pós-doutorado. O número de linhas associadas a cada docente era igual ao número de pós-doutorados feitos pelo(a) docente, de maneira que as instituições se repetiam. A instrução feita à IA foi que identificasse primeiro o país de origem de cada instituição com base em seu nome, contexto geográfico implícito e, quando necessário, conhecimento sobre instituições conhecidas. Em seguida, com base no país identificado, deveria ser feita a classificação do pós-doutorado em Nacional (feito no Brasil) ou Internacional.

Para validar a classificação, selecionou-se aleatoriamente 20 pesquisadores(as). Em seus currículos *Lattes*, verificou-se o número de pós-doutorados e as instituições onde foram realizados. Então, essas informações foram comparadas com aquelas retornadas pela IA. O procedimento obteve acurácia de 100% nesse teste, porém é importante destacar que, apesar do resultado positivo, o sistema de inteligência artificial está sujeito a erros, e por isso é recomendável verificar informações em mais de uma fonte de IA (CNN, 2025). No entanto, no momento de realização deste estudo, a Grok foi a única IA gratuita (conhecida pela autora) capaz de realizar o procedimento de classificação com boa acurácia no conjunto de dados utilizado.

5 RESULTADOS DAS ANÁLISES DESCRITIVAS E INFERENCIAIS

Com as análises descritivas ou exploratórias, buscou-se uma melhor compreensão do conjunto de dados e das características dos(as) pesquisadores(as) bolsistas e não-bolsistas de produtividade. Para isso, foi feito o uso de gráficos de barras, barras empilhadas e diagramas de caixas (*boxplots*), além do cálculo de estatísticas descritivas. Nos gráficos do tipo *boxplot*, os *outliers* foram omitidos para algumas variáveis com o intuito de preservar a visualização das caixas, mas foram considerados nas análises inferenciais e nas modelagens (Capítulo 6).

Com a finalidade de verificar se há evidências de diferenças significativas entre os indicadores de produtividade de diferentes grupos, o teste de hipóteses foi aplicado. Os grupos foram definidos pela bolsa de produtividade (PQs e Não-PQs) e pelo sexo (feminino e masculino) - para esse segundo caso, inicialmente foi considerada a totalidade da base de dados (isto é, homens e mulheres PQ e Não-PQ); e depois, apenas os(as) pesquisadores(as) PQ, considerando cada um dos níveis de produtividade (1A, 1B, 1C, 1D e 2). Dessa forma, as variáveis de contagem foram comparadas entre grupos via o teste da mediana, com o nível de significância fixado em 5%. Como o teste teve por finalidade encontrar evidências de diferença de produtividade entre grupos, ele foi definido como bilateral, testando (para cada indicador ou variável) a hipótese nula de igualdade das medianas de produções contra a hipótese alternativa de que essas medianas são diferentes. A determinação do parecer foi feita com base em análises gráficas e, principalmente, na verificação da proporção de observações acima da mediana combinada (ver Capítulo 3.1.1), considerando cada par de amostras em comparação.

Com respeito ao teste qui-quadrado para independência, este foi aplicado para uma tabela de contingência contendo as variáveis sexo (feminino e masculino) e nível PQ (indicador de bolsista de produtividade), de forma a verificar se há evidências de associação ou dependência entre essas variáveis. O nível de significância foi fixado em 5%.

Como o objetivo do trabalho é verificar não só a existência de critérios que distingam PQs de Não-PQs, como investigar se há evidências de diferença de produtividade entre homens e mulheres, as análises descritivas e inferenciais foram realizadas

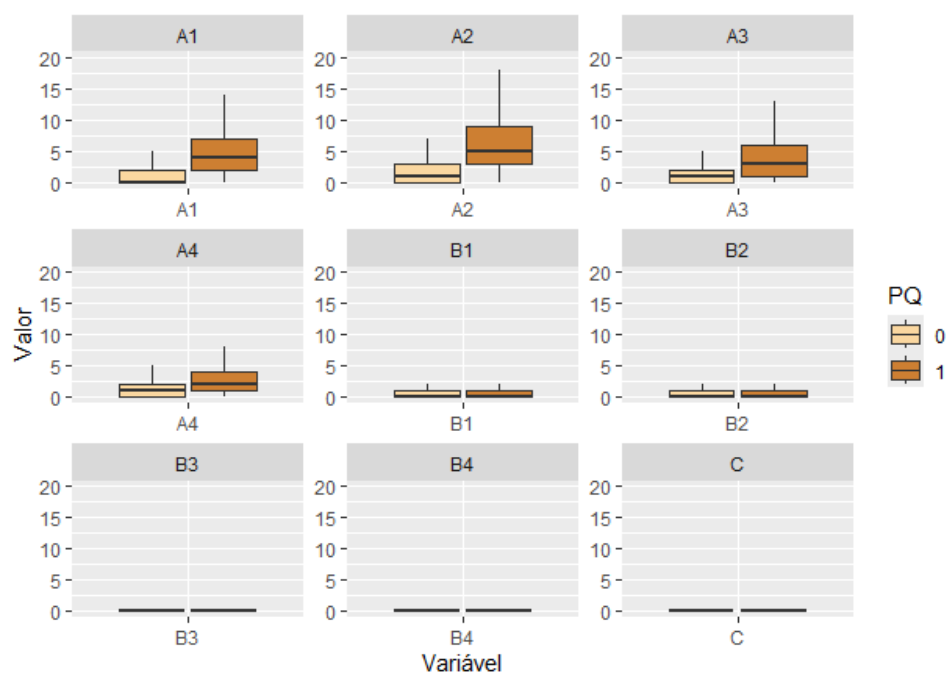
considerando dois pares de categorias distintas: PQs e Não-PQs, homens e mulheres. Dessa forma, o Capítulo 5.1 contempla as diferenças de produtividade encontradas entre pesquisadores(as) bolsistas e não-bolsistas; o Capítulo 5.2 investiga a diferença de produção entre sexos, considerando a totalidade dos(as) pesquisadores(as) (Capítulo 5.2.1), apenas os(as) pesquisadores(as) PQ (Seção 5.2.2) e os(as) pesquisadores(as) PQ por categoria/nível da bolsa (Capítulo 5.2.3).

5.1 PQs e Não-PQs

Dos(as) 1.383 pesquisadores(as) contabilizados (docentes dos PPGs da área “Matemática, Estatística e Probabilidade”), 427 são bolsistas de produtividade (31% do total de pesquisadores). Em contrapartida, 956 docentes são não-bolsistas (69%). Deixando de lado por um instante a questão do gênero, foram realizadas análises exploratórias e testes de hipóteses para verificar se há evidências de diferenças na produção científica entre PQs e Não-PQs. As variáveis (indicadores de produtividade) consideradas(os) foram as(os) relatadas(os) em Descrição e Processamento dos Dados (Capítulo 4).

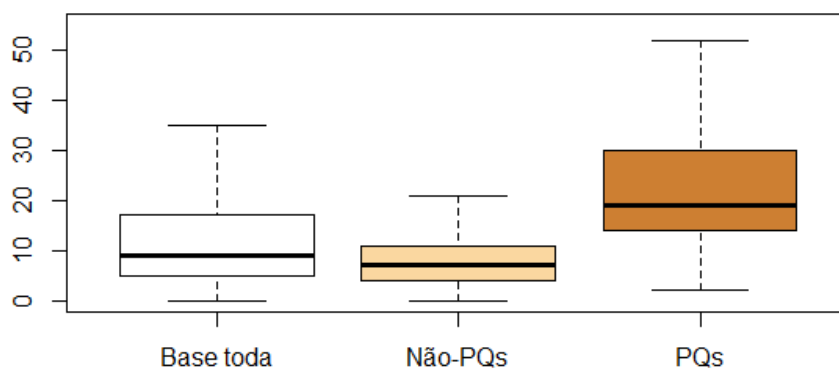
As Figuras 2 a 10 apresentam os diagramas de caixa construídos para cada variável de produtividade, considerando pesquisadores(as) bolsistas e não-bolsistas. Pela Figura 2 é possível perceber que a mediana da produção de artigos por PQs parece ser maior nas categorias A1, A2, A3 e A4. Além disso, a análise da Figura 3 sugere que a publicação total de artigos (mediana) também é maior para PQs. Outras diferenças claras para a mediana de produção ocorrem para: orientações de iniciações científicas concluídas, mestrados concluídos e doutorados em andamento e concluídos (Figura 7), com os(as) PQs orientando mais em todos os casos. Além disso, os(as) pesquisadores(as) bolsistas parecem ter mais anos de doutoramento (Figura 8), mais pós-doutorados feitos no exterior (Figura 9) e maior participação em eventos acadêmicos (Figura 10).

Figura 2: Quantidade de artigos publicados por categoria QUALIS. Zero (0) representa não-PQs enquanto (1) representa PQs.



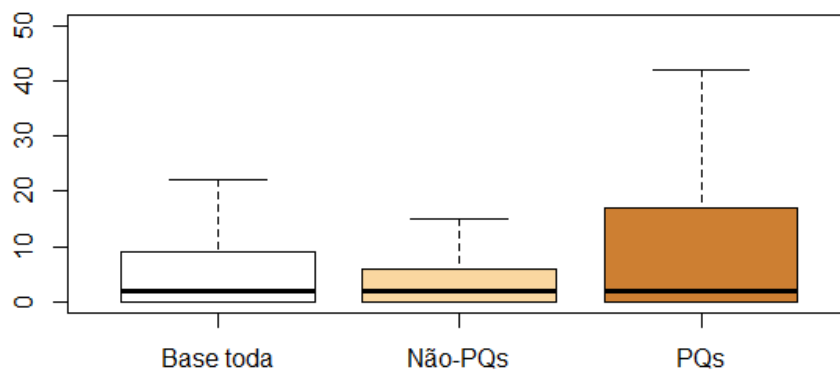
Fonte: autora (2025).

Figura 3: Total de artigos publicados (base toda, PQs e Não-PQs).



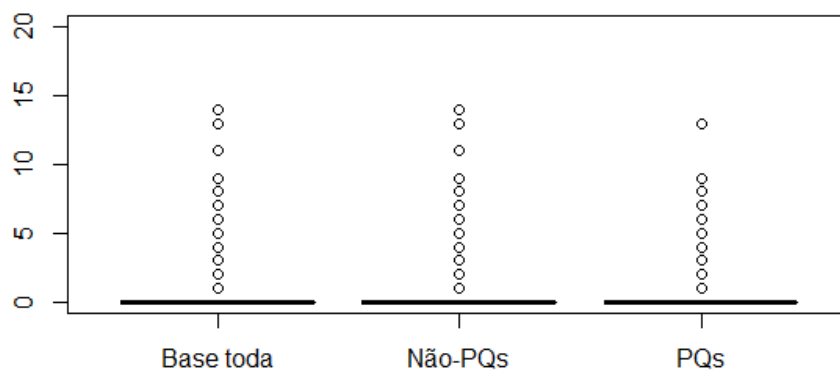
Fonte: autora (2025).

Figura 4: Publicações em congressos (base toda, PQs e Não-PQs).



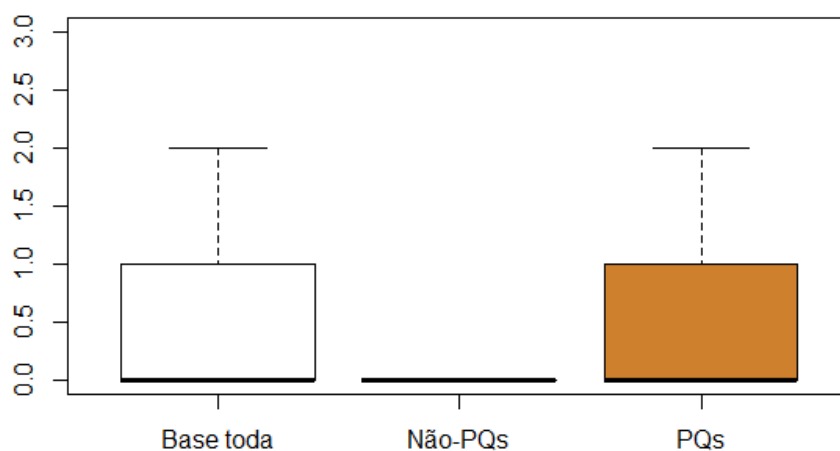
Fonte: autora (2025).

Figura 5: Livros publicados (base toda, PQs e Não-PQs).



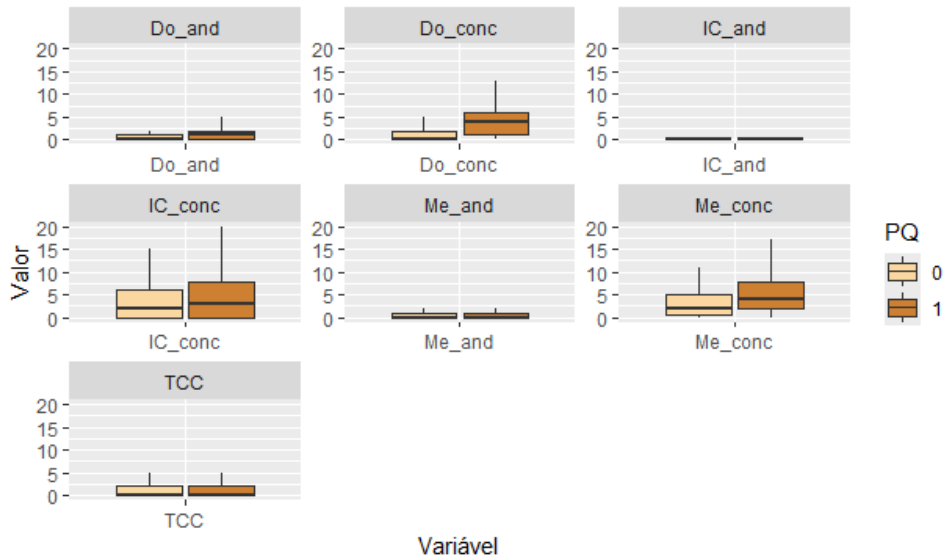
Fonte: autora (2025).

Figura 6: Capítulos de livros publicados (base toda, PQs e Não-PQs).



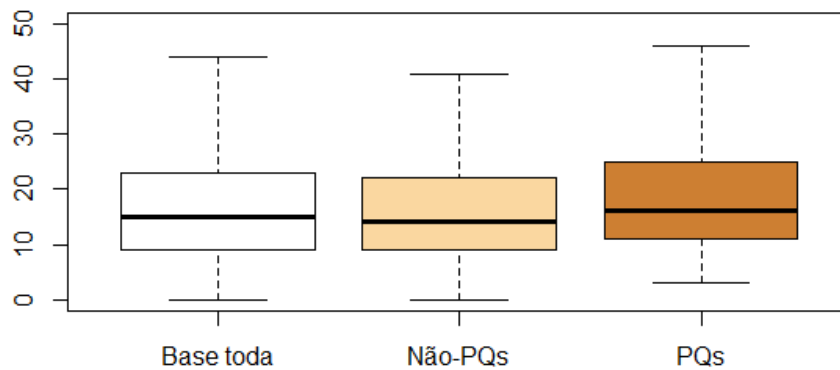
Fonte: autora (2025).

Figura 7: Orientações por categoria. Zero (0) representa não-PQs enquanto (1) representa PQs.



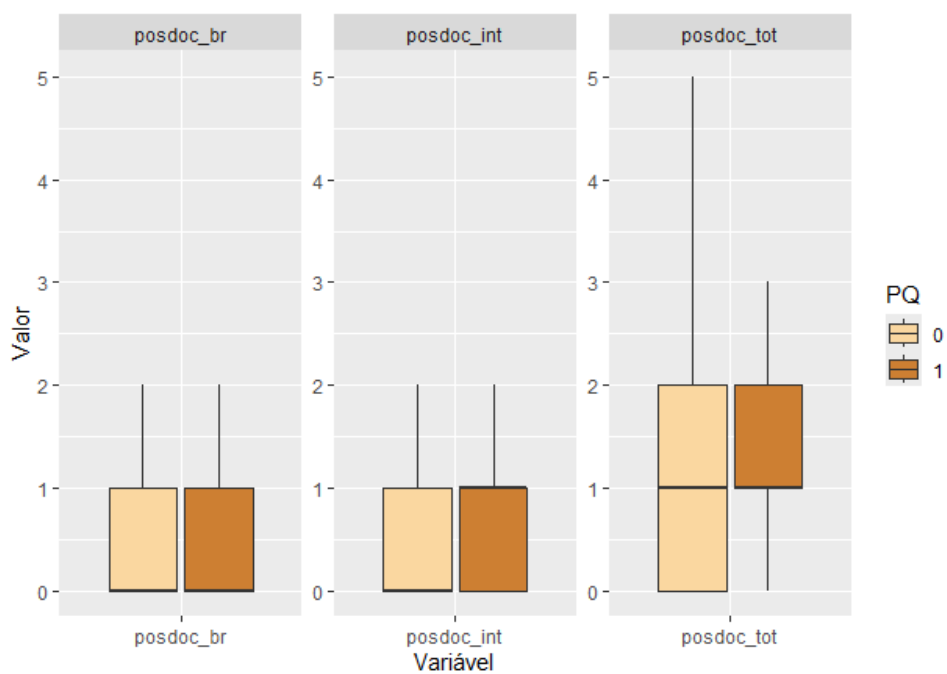
Fonte: autora (2025).

Figura 8: Anos desde doutoramento (PQs e Não-PQs).



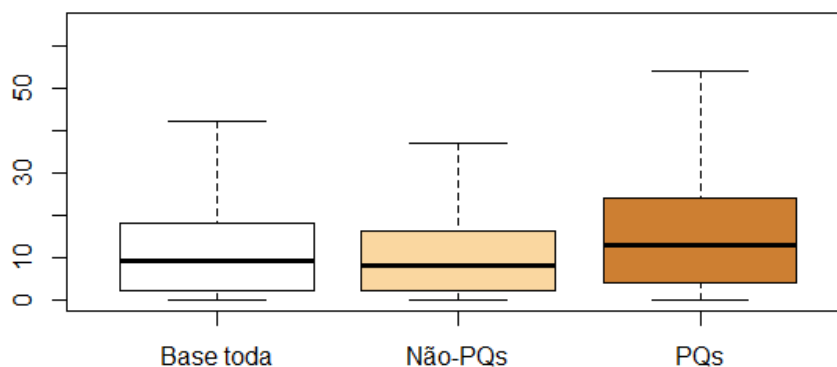
Fonte: autora (2025).

Figura 9: Pós doutorados. Zero (0) representa não-PQs enquanto (1) representa PQs.



Fonte: autora (2025).

Figura 10: Participações em eventos (PQs e Não-PQs).



Fonte: autora (2025).

Com o intuito de verificar se há evidências de diferenças significativas de produção (considerando cada variável de produtividade) entre PQs e Não-PQs, o teste de hipóteses para a mediana foi aplicado. Assim, a Tabela 3 apresenta a relação dos indicadores que se mostraram estatisticamente diferentes entre os grupos. Nessa tabela também são listadas as figuras associadas a cada indicador, os respectivos valores de p associados ao teste da mediana e os pareceres. Além disso, no Anexo B é possível conferir uma tabela contendo as estatísticas para cada indicador (Tabela B.2) e outra tabela contendo a relação dos indicadores que não se mostraram significativos (Tabela B.1).

Tabela 3: Indicadores estatisticamente (5%) diferentes entre PQs e Não-PQs.

Figura B	Indicador	P-Valor (mediana)	Parecer
2	Artigo A1	0,0000*	PQs publicam mais
2	Artigo A2	0,0000*	PQs publicam mais
2	Artigo A3	0,0000*	PQs publicam mais
2	Artigo A4	0,0000*	PQs publicam mais
2	Artigo B1	0,0092*	PQs publicam mais
2	Artigo B2	0,0009*	PQs publicam mais
2	Artigo B3	0,0518*	PQs publicam mais
3	Total de Artigos	0,0000*	PQs publicam mais
5	Livros	0,0000*	PQs publicam mais
6	Cap. Livros	0,0000*	PQs publicam mais
7	IC conc	0,0039*	PQs orientam mais
7	Mestrado conc	0,0000*	PQs orientam mais
7	Doutorado and	0,0000*	PQs orientam mais
7	Doutorado conc	0,0000*	PQs orientam mais
8	Anos de Doutorado	0,0003*	PQs possuem mais
9	Total de Pós Docs	0,0019*	PQs possuem mais
9	Pós Docs Exterior	0,0000*	PQs possuem mais
10	Eventos	0,0000*	PQs participam mais

Fonte: autora (2025).

Assim, considerando os indicadores que se mostraram significativos (Tabela 3) as Figuras 2 a 10 e as estatísticas da Tabela B.2 (Apêndice B), os(as) pesquisado-res(as) PQ apresentam mediana maior para: publicação de artigos A1, A2, A3, A4, B1, B2 e B3 (respectivamente); publicação total de artigos; livros e capítulos de li-vros publicados; orientações para ICs concluídas, mestrados concluídos e doutorados (em andamento e concluídos); anos desde o doutoramento, pós doutorados feitos no exterior, total de pós doutorados e número de eventos atendidos. Nenhum indicador apresentou mediana maior para os não-PQs.

Esses resultados indicam que, de fato, existem evidências que apontam para a existência de atributos objetivos que distinguem pesquisadores(as) bolsistas de pes-

quisadores(as) não-bolsistas de produtividade. Além disso, demonstram claramente como as bolsas são uma valorização das métricas acadêmicas dos(as) pesquisadores(as). Lembrando que os atributos (variáveis de produtividade) em estudo estão relacionados aos critérios de julgamento 1 (produção científica) e 3 (formação de recursos humanos) relatados na Tabela 1. Agora, como esses atributos se distinguem entre cientistas homens e mulheres, e como a distribuição das bolsas de produtividade se relaciona com o sexo do(a) pesquisadora(a), são aspectos a serem investigados nas seções subsequentes.

5.2 Status PQ e Sexo

Conforme dito em *Descrição e Pré-Processamento dos Dados* (Capítulo 4), os(as) pesquisadores(as) que integram o conjunto de dados utilizado neste estudo são da área de Estatística, Matemática e Probabilidade. Ao total, foram contabilizados(as) 1.383 pesquisadores(as), dos(as) quais 282 são mulheres (20%) e 1.101 são homens (80%), conforme mostra a Tabela 4. O percentual baixo de mulheres é, infelizmente, persistente em áreas de STEM, e ocorre por questões estruturais e culturais, conforme discutido no Capítulo 1.2.

Tabela 4: Distribuição de Homens, Mulheres e Total na base de dados.

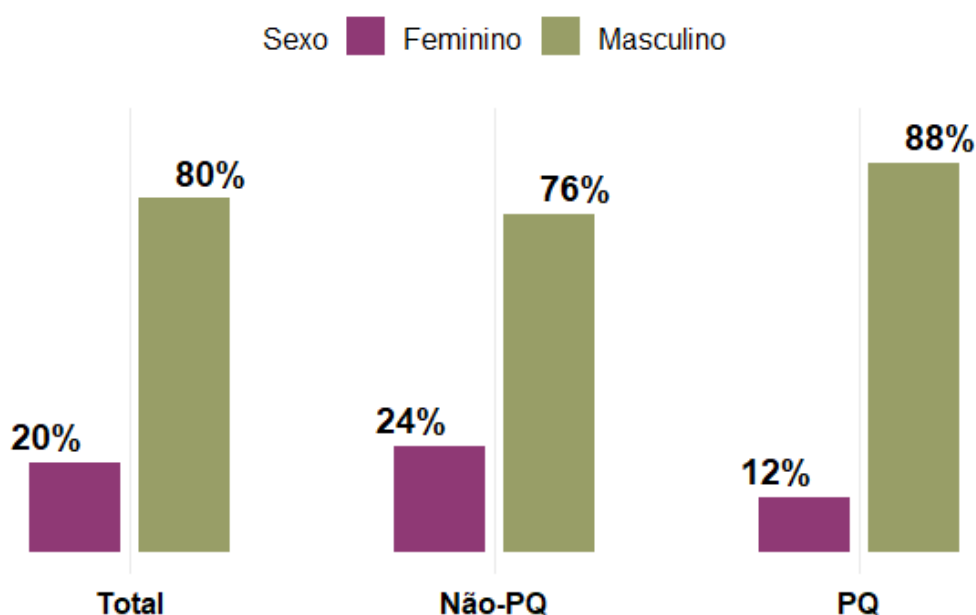
Sexo	Frequência	Percentual (%)
Homem	1101	80
Mulher	282	20
Total	1383	100

Fonte: autora (2025).

Na Figura 11, é possível conferir um gráfico que mostra as proporções de homens e mulheres dentro de cada um dos seguintes grupos: a totalidade dos(as) pesquisadores(as) da base (isto, PQs e Não-PQs combinados(as)), os(as) não-bolsistas e os(as) bolsistas de produtividade. Por esse gráfico, é possível perceber a manifestação do chamado efeito tesoura, isto é, uma redução do contingente feminino em grupos de maior prestígio na comunidade científica (Nery, 2025) - neste caso, no grupo dos(as) pesquisadores(as) bolsistas de produtividade. Pela imagem (Figura 11), percebe-

se que as mulheres representam 20% da totalidade dos(as) pesquisadores(as), 24% dos(as) não-bolsistas e apenas 12% dos(as) bolsistas. Por outro lado, os homens compreendem 80% dos(as) pesquisadores(as) da base, 76% dos não-bolsistas e expressivos 88% dos(as) pesquisadores(as) bolsistas de produtividade.

Figura 11: Distribuição de Pesquisadores(as) por Sexo e Status PQ.



Fonte: autora (2025).

A Tabela 5 mostra o quantitativo de PQs e não-PQs, segregados por sexo. Considerando os dados apresentados nessa tabela, foi aplicado o teste qui-quadrado de independência para avaliar a associação entre sexo e recebimento de bolsa de produtividade. O resultado indicou uma associação estatisticamente significativa entre as variáveis ($\chi^2 = 23,52$, $GL = 1$, $p\text{-valor} < 0,05$), sugerindo que o recebimento de bolsas PQ não é independente do sexo. Uma razão de chances (*odds*) que descreve essa desigualdade é $(374/727)/(53/229) = 2,22$. Ou seja, dentre os(as) pesquisadores(as) da área “Estatística, Matemática e Probabilidade”, a chance estimada de ser bolsista de produtividade é 122% maior para os homens em comparação com as mulheres (e com aquilo que seria esperado sob a hipótese de independência entre sexo e recebimento da bolsa PQ).

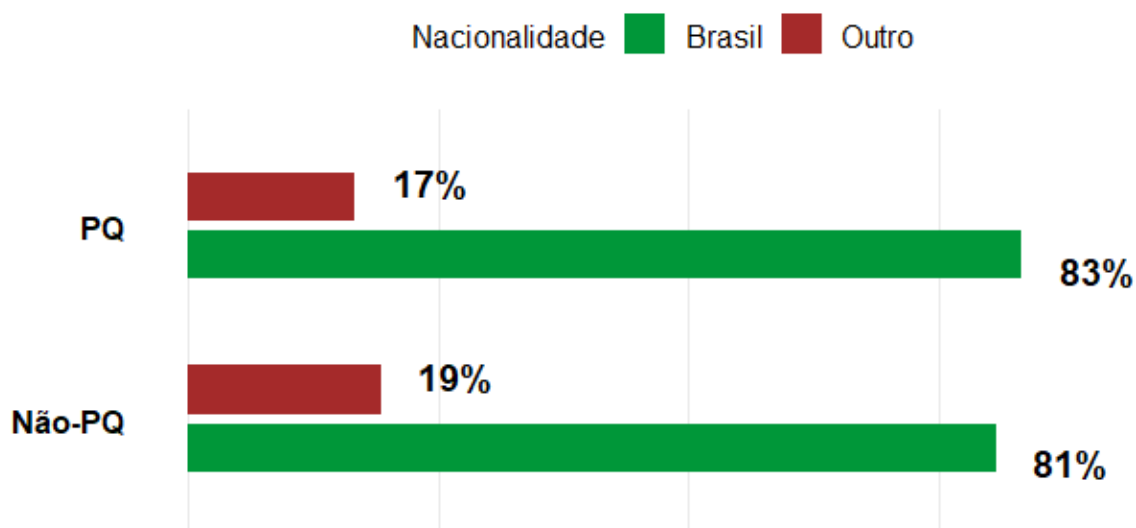
Tabela 5: Distribuição de bolsas de produtividade por sexo.

	Homens	Mulheres	Total
PQ	374	53	427
Não-PQ	727	229	956
Total	1106	282	1383

Fonte: autora.

Na Figura 12 é possível conferir a proporção de pesquisadores(as) PQ e Não-PQ por país de origem (Brasil ou Outro). Os percentuais não se alteram muito por status PQ: os(as) nascidos(as) no Brasil são 81% dos(as) não-bolsistas e 83% dos(as) bolsistas, um incremento de apenas 2 pontos percentuais. De forma complementar, os(as) pesquisadores(as) nascidos no exterior integram 19% dos(as) não-bolsistas e 17% dos(as) bolsistas. Neste estudo, verificou-se a existência de 42 nacionalidades distintas entre os(as) pesquisadores(as), sendo a brasileira a mais frequente (nacionalidade de 1.141 ou 83% dos(as) pesquisadores(as)).

Figura 12: Distribuição de Pesquisadores(as) PQ e Não-PQ por País de Origem.

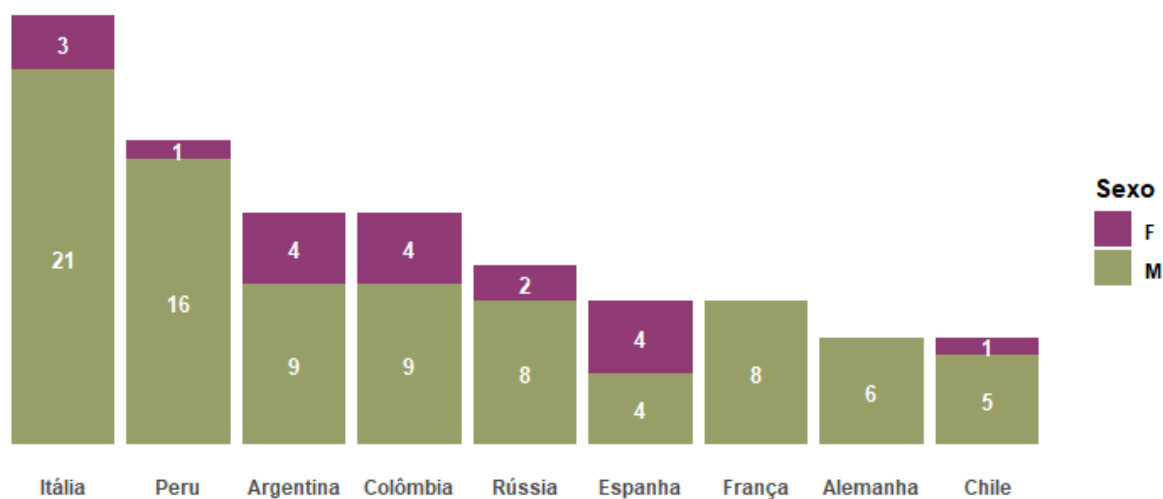


Fonte: autora (2025).

As Figuras 13 e 14 são gráficos de barras empilhadas para a proporção de mulheres e homens estrangeiros(as), dentre não-bolsistas e bolsistas de produtividade e

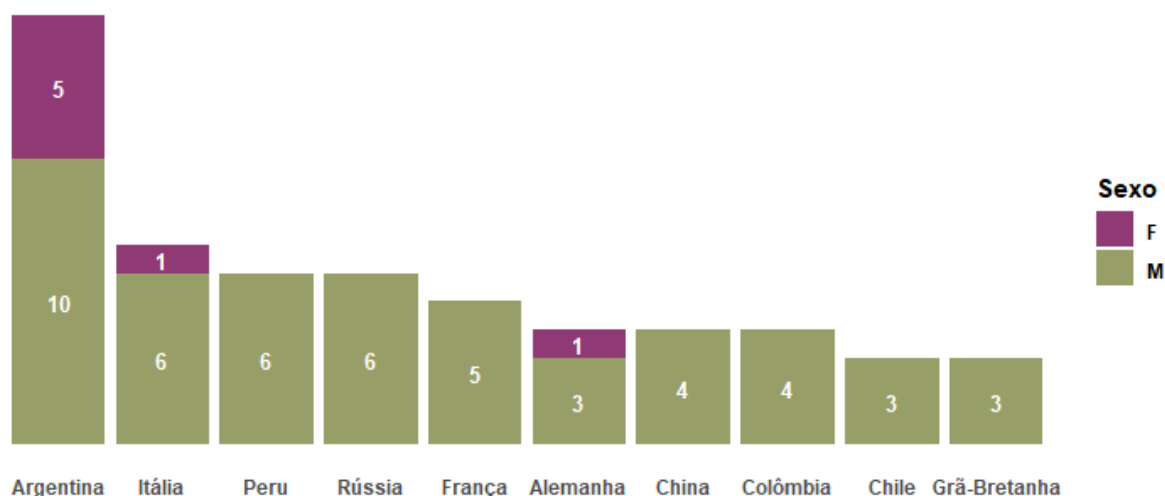
considerando os 10 países estrangeiros mais frequentes entre todos(as) os(as) pesquisadores(as). Esses gráficos revelam que, com exceção da Espanha no grupo dos não-PQs, as mulheres também estão sub-representadas entre os(as) estrangeiros(as). A situação é ainda mais preocupante se considerarmos apenas o grupo de pesquisadores(as) PQ, no qual 7 dos 10 países estrangeiros apresentam somente homens como representantes (as mulheres, com exceção da França, da China e da Grã-Bretanha, que não possuem nenhuma, são não-bolsistas).

Figura 13: Distribuição de Não-PQs por Sexo para os Top 10 Países Estrangeiros.



Fonte: autora (2025).

Figura 14: Distribuição de PQs por Sexo para os Top 10 Países Estrangeiros.



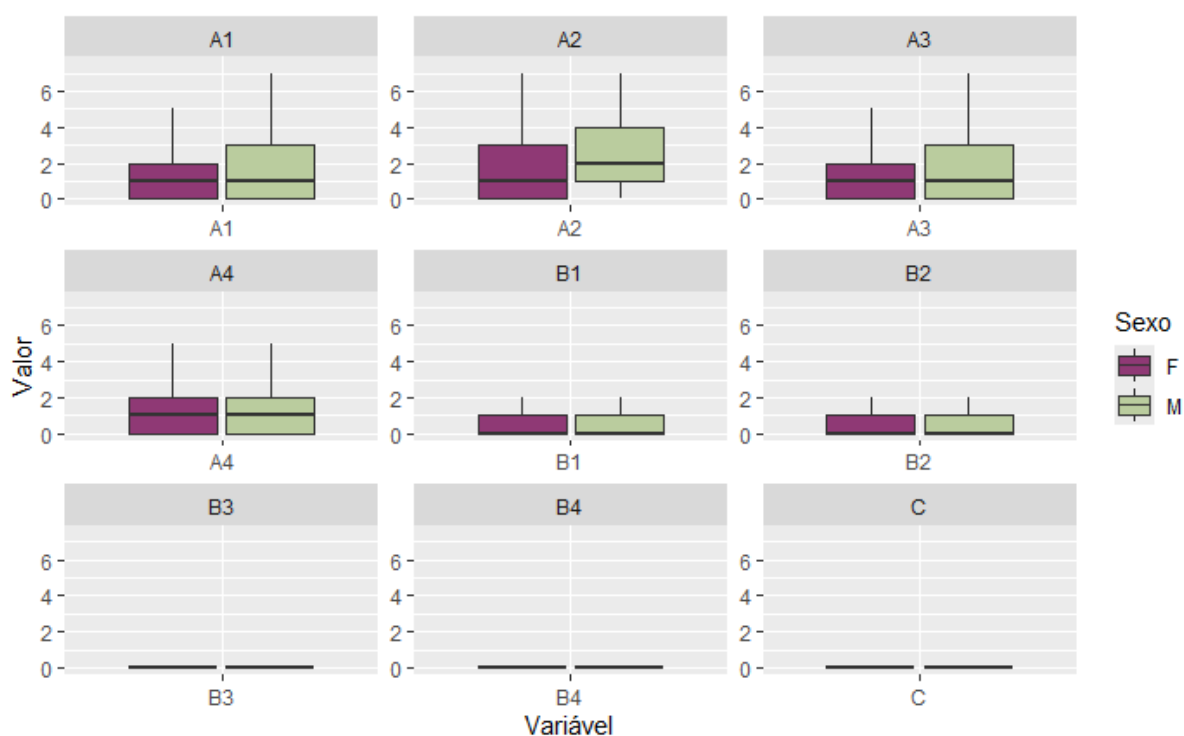
Fonte: autora (2025).

Considerando apenas os(as) pesquisadores(as) de origem brasileira, os resultados não são muito diferentes, conforme já foi discutido e apresentado na Figura 11. Esses dados demonstram que a desigualdade de gênero é um problema que ultrapassa barreiras culturais e geográficas, o que reforça a necessidade de investigações mais profundas sobre seus impactos, além da adoção de políticas que promovam sua mitigação.

5.2.1 Diferenças entre Sexos (Juntando PQs e Não-PQs)

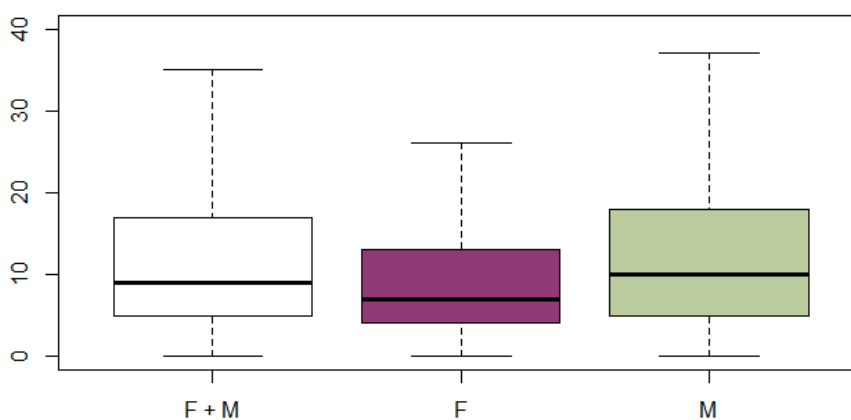
Para a análise exploratória da distribuição dos indicadores por sexo do(a) pesquisador(a), foram construídos os gráficos das Figuras 15 a 23. As estatísticas associadas a esses indicadores são apresentadas no Anexo C, Tabela C.2. Pela análise gráfica e considerando as medianas de cada distribuição, homens parecem publicar mais artigos A2 e artigos em geral; já as mulheres parecem apresentar maior participação na publicação de capítulos de livros, orientação para IC concluída e presença em eventos.

Figura 15: Quantidade de artigos publicados por categoria QUALIS (Sexos Feminino e Masculino).



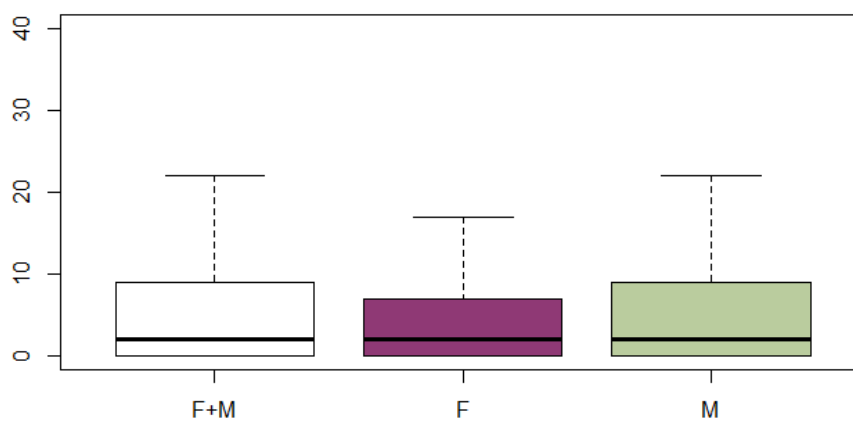
Fonte: autora (2025).

Figura 16: Total de artigos publicados (Base toda, Sexos Feminino e Masculino).



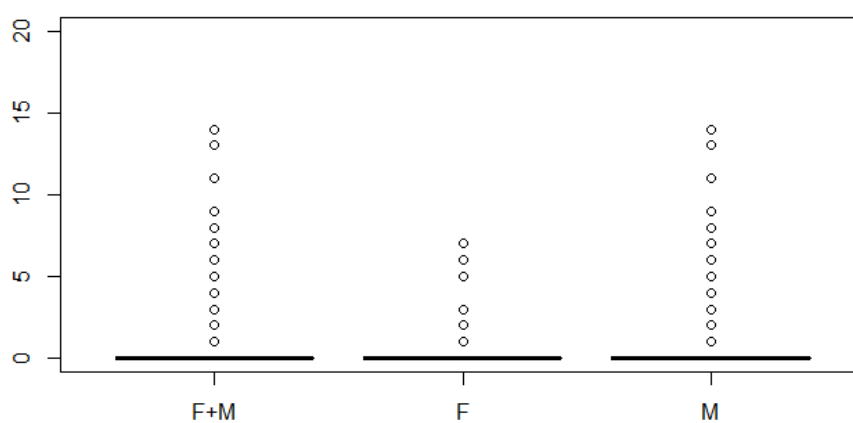
Fonte: autora (2025).

Figura 17: Publicações em Congresso (Base Toda, Sexos Feminino e Masculino).



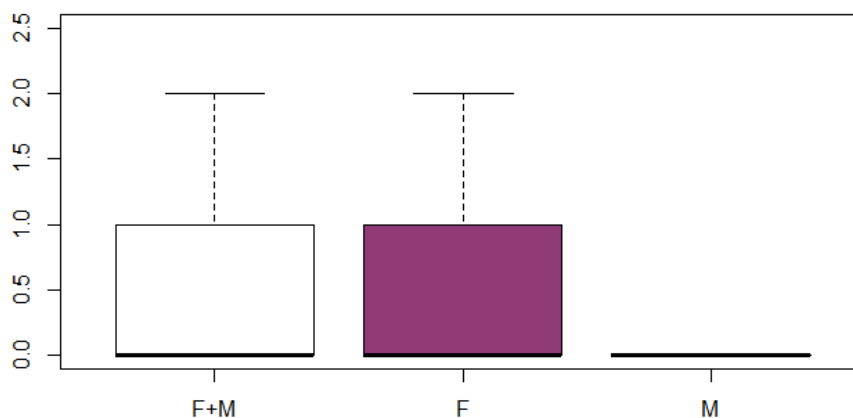
Fonte: autora (2025).

Figura 18: Número de livros publicados (Base Toda, Sexos Feminino e Masculino).



Fonte: autora (2025).

Figura 19: Número de capítulos de livros publicados (Base Toda, Sexos Feminino e Masculino).



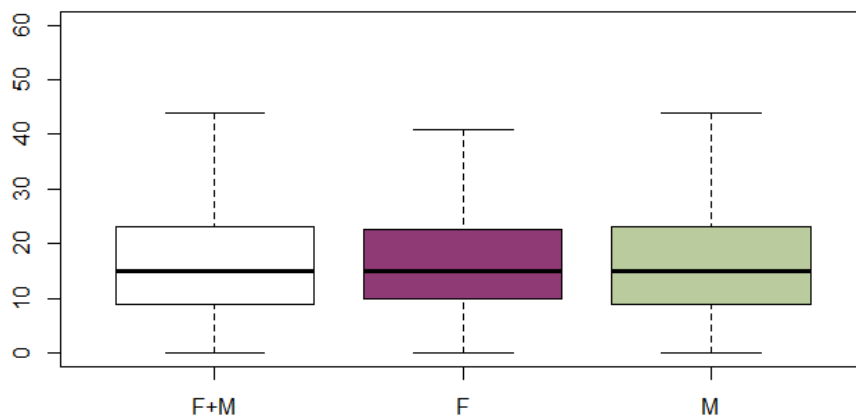
Fonte: autora (2025).

Figura 20: Orientações por categoria (Sexos Feminino e Masculino).



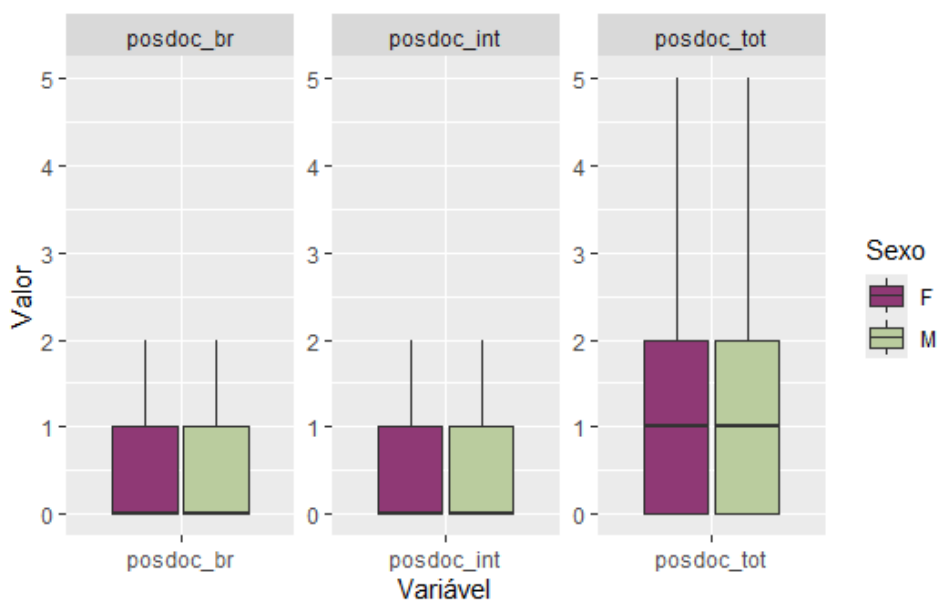
Fonte: autora (2025).

Figura 21: Anos desde o doutoramento (Sexos Feminino e Masculino).



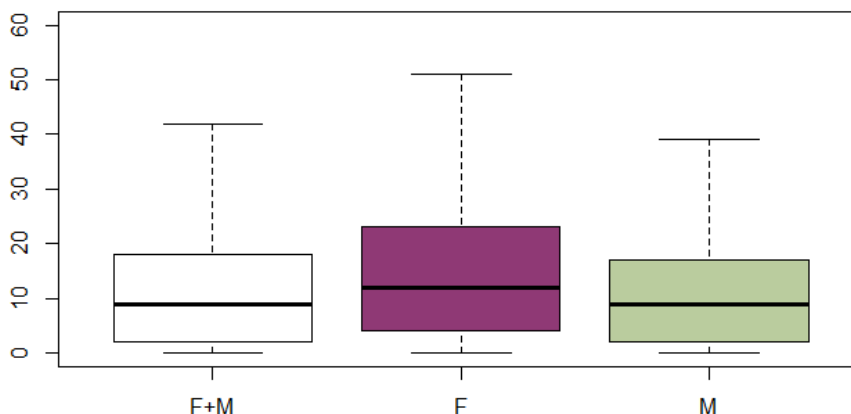
Fonte: autora (2025).

Figura 22: Quantidade de pós doutorados (Sexos Feminino e Masculino).



Fonte: autora (2025).

Figura 23: Participações em eventos (Sexos Feminino e Masculino).



Fonte: autora (2025).

Considerando a totalidade dos(as) pesquisadores(as) da base de dados - isto é, sem fazer distinção entre PQs e Não-PQs -, houve evidências de diferenças significativas (5%) entre sexos para os seguintes indicadores: quantidade de artigos escritos por categoria (A1, A2, A3 e A4), total de artigos escritos, quantidade de orientações (exceto para mestrado concluído), número de capítulos de livros publicados e número de eventos atendidos. A relação dos indicadores significativos por figura, p-valor e parecer pode ser conferida na Tabela 6. Ademais, a Tabela C.1, contida no Anexo C, apresenta a relação dos indicadores que se mostraram estatisticamente iguais entre homens e mulheres.

Tabela 6: Indicadores estatisticamente (5%) diferentes entre homens e mulheres (base toda).

Figura	Indicador	Valor-p (mediana)	Parecer
15	Artigo A1	0,0009*	Homens publicam mais
15	Artigo A2	0,0000*	Homens publicam mais
15	Artigo A3	0,0000*	Homens publicam mais
15	Artigo A4	0,0000*	Homens publicam mais
16	Total de Artigos	0,0000*	Homens publicam mais
20	Doutorado and	0,0250*	Homens orientam mais
20	Doutorado conc	0,0014*	Homens orientam mais
20	IC and	0,0041*	Mulheres orientam mais
20	IC conc	0,0000*	Mulheres orientam mais
20	Mestrado and	0,0192*	Mulheres orientam mais
20	TCC	0,0000*	Mulheres orientam mais
19	Cap. Livros	0,0000*	Mulheres publicam mais
23	Eventos	0,0144*	Mulheres participam mais

Fonte: autora.

Os resultados (p-valores para o teste da mediana) apresentados na Tabela 6, juntamente às estatísticas apresentadas na Tabela C.2 (Anexo C), evidenciam que homens publicam mais artigos (ao total e por categoria A1, A2, A3, A4) e orientam mais para doutorado (em andamento e concluído). Já as mulheres orientam mais para IC (em andamento e concluída), TCC e mestrado (em andamento), além de publicarem mais capítulos de livros e participarem mais de eventos - agora, se tivéssemos a infor-

mação de quais eventos são nacionais e quais são internacionais, talvez a participação feminina reduzisse nos internacionais devido à existência de fenômenos como a dupla jornada de trabalho, que afeta mais as mulheres (PUCPR Digital, 2023). Além disso, é importante salientar que os dados disponíveis no *Lattes* refletem as informações que o(a) pesquisador(a) optou por divulgar (ou se recordou de incluir).

Os achados relatados na Tabela 6 sustentam os resultados de Lopes (2025) sobre os homens publicarem e orientarem nos níveis de maior prestígio, enquanto as mulheres orientam mais nos níveis menos prestigiados (IC e TCC). Isso acende o alerta para os efeitos que essas diferenças podem ter em estágios mais avançados da carreira dos(as) pesquisadores(as), com as diferenças entre homens e mulheres tendendo a se acentuar (Lopes, 2025).

5.2.2 Diferenças entre Sexos (Apenas PQs)

Considerando, agora, o grupo de pesquisadores(as) composto somente pelos(as) bolsistas de produtividade, o número de diferenças significativas (5%) entre sexos reduziu em relação ao caso geral (amostra composta por bolsistas e não-bolsistas), sendo essas diferenças encontradas apenas para as seguintes variáveis: publicação de artigos A3, orientações para IC concluída, TCC e doutorado concluído, além do número de capítulos de livros publicados. A Tabela 7 apresenta a relação dos indicadores estatisticamente diferentes entre sexos por p-valor do teste e parecer. Os pareceres foram estabelecidos com base na proporção de pontos acima da mediana combinada, considerando grupos definidos pelo sexo (feminino e masculino). Essas estatísticas podem ser conferidas na Tabela C.3, contida no Apêndice C.

A partir deste ponto, serão apresentadas apenas as estatísticas referentes aos indicadores significativos. Optou-se por omitir os gráficos e medidas descritivas completas para todos os indicadores analisados a fim de evitar repetições excessivas e sobrecarga de conteúdo. Ressalta-se, contudo, que todos os indicadores — incluídos ou não nas visualizações — foram tratados com os mesmos procedimentos metodológicos empregados até aqui e descritos em *Descrição e Pré-Processamento dos Dados*, garantindo a uniformidade e a comparabilidade das análises realizadas.

Tabela 7: Indicadores estatisticamente (5%) diferentes entre homens e mulheres (PQs).

Indicador	P-Valor	Parecer
Artigo A3	0,0136*	Homens publicam mais
IC conc	0,0049*	Mulheres orientam mais
Doutorado conc	0,0483*	Mulheres orientam mais
TCC	0,0148*	Mulheres orientam mais
Cap. Livros	0,0017*	Mulheres publicam mais

Fonte: autora.

Ao analisarmos de maneira conjunta os dois cenários (caso geral e somente PQs), percebemos a prevalência de algumas conclusões: homens publicam mais artigos A3, mulheres orientam mais alunos de graduação (IC e TCC) e publicam mais capítulos de livros. Um resultado relevante é que, apesar de no caso geral parecer que os homens orientaram mais doutorados, no grupo composto apenas por PQs acontece o contrário: as evidências indicam que as mulheres orientaram mais doutorados.

Os resultados da Tabela 7 sugerem que cientistas mulheres necessitam de maiores realizações para alcançar o mesmo nível de reconhecimento que cientistas homens. Isso aponta para a possível existência de vieses embutidos nos processos de avaliação e atribuição de mérito na comunidade científica, nos quais o trabalho e as contribuições de mulheres podem ser subvalorizados em comparação com os de homens. Para uma compreensão mais aprofundada dessa dinâmica no contexto deste estudo, seria crucial ter acesso aos critérios específicos utilizados para a distribuição das bolsas de produtividade em pesquisa, bem como analisar a composição dos comitês de avaliação e as potenciais influências de estereótipos de gênero em suas decisões. Adicionalmente, pesquisas futuras poderiam explorar as trajetórias de carreira de cientistas mulheres e homens, comparando a obtenção de financiamento para pesquisa, o recebimento de prêmios e honrarias, e sua representação em cargos de liderança, buscando quantificar e qualificar as possíveis barreiras que as mulhe-

res enfrentam para obter reconhecimento equivalente ao de seus pares masculinos. Compreender as causas subjacentes a essa disparidade é fundamental para o desenvolvimento de estratégias eficazes que promovam a equidade de gênero e garantam que o talento e as contribuições de todos os(as) cientistas sejam devidamente reconhecidos e valorizados.

5.2.3 *Diferenças entre Sexos por Categoria e Nível PQ*

Considerando o grupo de pesquisadores(as) bolsistas de produtividade, a Tabela 8 apresenta a distribuição de mulheres (F) e homens (H) por categoria e nível PQ. Com isso, percebe-se que as mulheres estão subrepresentadas em todas as categorias/níveis - em especial, na categoria/nível de maior prestígio (1A), onde 90% dos(as) integrantes são homens e apenas 10% são mulheres. De agora em diante, os termos “categoria” e “nível” serão utilizados de forma intercambiável, referindo-se ao mesmo conceito (isto é, os grupos de produtividade 1A, 1B, 1C, 1D e 2).

Tabela 8: Distribuição de bolsistas por nível e gênero.

Categoria/Nível	F	M	Total	% F	% M
1A	3	26	29	10	90
1B	5	28	33	15	85
1C	7	32	39	18	82
1D	13	127	140	9	91
2	25	161	186	13	87

Fonte: autora (2025).

Considerando cada categoria PQ de pesquisador(a), foram encontradas diferenças significativas (5%) de produtividade entre mulheres e homens em 1D (número de pós doutorados e participação em eventos) e 2 (capítulos de livros publicados e orientações para doutorado concluído) - a relação dos indicadores que se mostraram estatisticamente diferentes entre sexos por categoria, p-valor e parecer pode ser conferida na Tabela 9. Além disso, as estatísticas descritivas para a mediana e mediana combinada (além da proporção de pontos acima da mediana combinada, considerando cada grupo definido pelo sexo) podem ser conferidas na Tabela C.4, contida no Apên-

dice C. Para as três categorias de maior prestígio (1A, 1B e 1C), não foram encontradas diferenças significativas entre sexos para nenhum dos indicadores considerados.

Tabela 9: Indicadores estatisticamente diferentes (5%) entre homens e mulheres por categoria PQ.

Categoria/Nível	Indicador	Valor-p (mediana)	Parecer
1D	Pós Doc Total	0,0126*	Mulheres possuem mais
1D	Eventos	0,0492*	Mulheres participam mais
2	Cap. Livros	0,0005*	Mulheres publicam mais
2	Doutorado conc	0,0049*	Mulheres orientam mais

Fonte: autora.

Os resultados apresentados na Tabela 9 reforçam a argumentação desenvolvida no Capítulo 5.2.2, sobre as cientistas mulheres precisarem demonstrar uma produção mais robusta para alcançar o mesmo reconhecimento que seus pares homens. Observa-se que, mesmo quando comparadas a homens na mesma categoria de produtividade (1D e 2), as mulheres apresentam desempenho superior para quatro indicadores — número de pós-doutorados, participações em eventos, capítulos de livros publicados e orientações para doutorado concluído. Esses achados respaldam o pensamento de que existem assimetrias de gênero no processo de avaliação da produtividade científica, indicando que as mulheres, para alcançarem a mesma posição que os homens, podem estar sendo cobradas por um volume maior de realizações acadêmicas.

6 RESULTADOS DA MODELAGEM ESTATÍSTICA

Com base nos resultados dos testes aplicados no Capítulo 5, foram ajustados dois modelos para verificar se, com base nas diferenças encontradas, é possível (1) classificar pesquisadores(as) em não-bolsistas e bolsistas (Capítulo 6.1) e (2), dado que o(a) pesquisador(a) é bolsista, identificar a categoria/nível de produtividade (Capítulo 6.2). Por fim, realizou-se a interpretação em termos de gênero e produtividade para as estimativas dos parâmetros dos modelos ajustados.

6.1 Ajuste do Modelo Logístico Binário (PQs e Não-PQs)

O primeiro modelo ajustado teve por objetivo distinguir pesquisadores(as) bolsistas de não-bolsistas. Nesse caso, como a variável resposta é dicotômica (PQ ou Não-PQ) e existe independência entre as observações (pesquisadores distintos), adotou-se o modelo logístico binário.

Como o conjunto de dados utilizado neste estudo continha muitas variáveis, com algumas sendo combinações lineares de outras, adotou-se um procedimento de pré-seleção de variáveis com base nos p-valores obtidos pelo teste da mediana (descrito no Capítulo 3.1.2), de acordo com os seguintes critérios:

1. Variáveis que apresentaram p-valor $< 10\%$ foram selecionadas;
2. Variáveis que são combinações lineares de outras (dentre as que passaram pela etapa 1) não foram selecionadas (é o caso para: total de artigos publicados).

Como as variáveis referentes à quantidade de pós-doutorados realizados no Brasil e no exterior (respectivamente) não foram pré-selecionadas — por terem apresentado p-valor maior que 10% no teste da mediana — optou-se pela pré-seleção da variável que representa o total de pós-doutorados, embora essa variável seja uma combinação linear (soma) dos números de pós-doutorados nacionais e internacionais. Ademais, a variável “nacionalidade” foi desconsiderada pela proporção de estrangeiros ser baixa (17%) e por ser composta por várias categorias (países) com representantes únicos (amostras muito pequenas).

Dessa forma, foram pré-selecionadas as seguintes variáveis: sexo, quantidade de publicações de artigos A1, A2, A3, A4, B1, B2 e B3 (respectivamente); publica-

ções em congressos, número de livros e de capítulos de livros publicados; orientações para IC concluída, mestrado concluído, doutorado em andamento e concluído, além de TCC; total de pós-doutorados, participações em eventos e anos desde o doutoramento.

Em seguida, empregou-se uma técnica de divisão do banco de dados na proporção de 80 para 20, com o primeiro conjunto sendo separado para treinamento do modelo e o segundo, para teste (James et al, 2021). Para o ajuste do modelo, a função do R utilizada foi a *glm* com o argumento *family = binomial* (R Core Team, 2024). Após o ajuste, realizou-se a retirada das variáveis não-significativas ($> 5\%$) por ordem decrescente do p-valor, com exceção da variável “sexo” e deixando o intercepto a ser retirado por último caso não se mostrasse significativo. Dessa forma, obteve-se um modelo com as seguintes variáveis: número de artigos A1, A2 e A3 publicados, capítulos de livros publicados, quantidade de orientações para mestrado (concluído) e doutorado (em andamento e concluído), além do número de eventos atendidos. Com exceção da variável sexo, todas as demais se mostraram significativas no modelo (considerando o nível de significância de 5% fixado).

Posteriormente, foi ajustado um novo modelo com as variáveis restantes (no modelo descrito no parágrafo anterior) e suas interações com “sexo”, pois desejamos avaliar se há algum viés de gênero embutido no ajuste (caso as interações não sejam significativas, a variável sexo será excluída da modelagem ao final). Então, realizou-se novamente o procedimento de retirada das variáveis/interações não-significativas por ordem decrescente do p-valor e deixando o intercepto a ser retirado por último caso não apresentasse significância (5%). Dessa forma, o modelo resultante ficou com as seguintes variáveis: sexo, número de artigos A1, A2 e A3 publicados, capítulos de livros publicados, quantidade de orientações para mestrado concluído, doutorado em andamento e concluído; quantidade de eventos atendidos, além das interações de sexo com orientações para mestrado concluído e doutorado concluído.

6.1.1 Classificação e Análise de Produtividade (PQs e Não-PQs)

Considerando o modelo logístico binário ajustado, a Tabela 10 apresenta a relação das variáveis explicativas (e do intercepto) por estimativa do coeficiente, erro padrão e p-valor.

Tabela 10: Especificação da regressão logística binária ajustada.

Variável	Estimativa	Erro Padrão	P-Valor
(Intercepto)	-3,5972	0,2260	0,0000*
Sexo	-0,3356	0,4383	0,4439
A1	0,2571	0,0318	0,0000*
A2	0,2321	0,0305	0,0000*
A3	0,1411	0,0336	0,0000*
Cap. Livros	-0,0780	0,0340	0,0218*
Me_conc	0,0679	0,0248	0,0063*
Do_and	0,1731	0,0716	0,0157*
Do_conc	0,1136	0,0389	0,0035*
eventos	0,0147	0,0061	0,0170*
Me_conc_Sexo	-0,2037	0,0939	0,0302*
Do_conc_Sexo	0,2910	0,1154	0,0117*

Fonte: autora.

Com o modelo especificado, foi feita a interpretação para as estimativas dos parâmetros (associados às variáveis explicativas ou indicadores de produtividade, além das interações com a variável sexo). Assim, para cada artigo publicado em periódicos A1, A2 e A3, a chance (*odds*) de um(a) pesquisador(a) ser classificado(a) como PQ aumenta em 29%, 26% e 15%, respectivamente. Importante notar que essas publicações são características dos homens, conforme evidenciado nas Tabelas 6 e 7.

Para cada orientação de doutorado em andamento, a chance de ser classificado(a) como bolsista aumenta em 19%, e para cada participação em evento acadêmico, a chance aumenta em 1,5%. Já para cada capítulo de livro publicado, a chance de ser classificado(a) como PQ reduz em 8%. Esse último resultado merece atenção, pois as mulheres são justamente o grupo que mais publica capítulos de livros, tanto no conjunto de todos(as) os(as) pesquisadores(as) (Tabela 6) quanto apenas no grupo de pesquisadores(as) bolsistas (Tabela 9). Isso levanta questões sobre a adequação dos critérios utilizados para mensurar produtividade e excelência, uma vez que podem subvalorizar formatos de publicação mais comuns entre mulheres.

Além disso, a cada orientação para mestrado concluído, a chance de ser classi-

ficado(a) como PQ aumenta em 7% para os homens e reduz em 13% para as mulheres. Esse resultado é particularmente expressivo e preocupante, pois sugere que a mesma atividade acadêmica (orientar mestrado) é valorizada de forma oposta dependendo do sexo do(a) pesquisador(a). Isso reforça a hipótese de viés sistêmico ou estrutural na forma como o mérito é atribuído no sistema de bolsas de produtividade.

Por fim, a cada orientação para doutorado concluído, a chance de ser classificado(a) como bolsista incrementa em 12% para os homens e em 50% para mulheres - ou seja, a orientação de doutorado tem um impacto mais expressivo para mulheres na probabilidade de receber bolsa de produtividade. Ao mesmo tempo, outras atividades acadêmicas mais comuns entre mulheres (como publicação de capítulos de livros e orientação de IC ou TCC - ver Capítulos 5.2.1 e 5.2.2, Tabelas 6 e 7) não são igualmente valorizadas.

Concluindo, a depender do sexo, a mesma contribuição acadêmica é interpretada de forma distinta. Isso reforça a noção de que mulheres cientistas podem estar submetidas a critérios mais rigorosos de julgamento, precisando atingir níveis mais altos de desempenho (como número elevado de orientações de doutorado) para alcançar reconhecimento semelhante ao de seus pares homens.

Para a verificação do pressuposto de ausência de multicolinearidade entre as variáveis explicativas do modelo, foi realizada a análise exploratória da matriz de correlações e o cálculo do *Variance Inflation Factor* (VIF). De acordo com Field et al (2012), correlações acima de 0,8 e valores acima de 10 para o VIF podem indicar a existência de multicolinearidade entre variáveis. No entanto, as correlações entre as variáveis da Tabela 10 ficaram todas abaixo de 0,8. Além disso, os valores calculados para o VIF ficaram todos abaixo de 10. Logo, não há indícios de multicolinearidade entre as variáveis explicativas.

A seguir, na Tabela 11, são apresentados os resultados da classificação feita para os(as) pesquisadores(as) da base-teste. Para essa previsão, a acurácia obtida foi de 83%, com IC 95% = (0,78; 0,88). A classe “positiva” foi definida como a classe dos(as) bolsistas de produtividade. Assim, a sensibilidade obtida (isto é, a proporção de PQs corretamente classificados(as) dessa maneira) foi de 59% e a especificidade (proporção de não-PQs corretamente classificados(as)) foi de 94%. Já a acurácia balanceada (média entre a sensibilidade e a especificidade) foi de 76,5%. Por fim, o *f1*

score para a classificação feita foi de 68%, indicando que a taxa de acerto do modelo para os(as) pesquisadores(as) classificados(as) como PQ é boa, mas ele está deixando de classificar corretamente uma quantidade razoável de bolsistas. A acurácia balanceada, a sensibilidade, a especificidade e a *f1 score* são estatísticas cuja análise é indicada em contextos com desbalanceamento das classes da variável resposta (Izbicki e Santos, 2020).

Tabela 11: Resultados da classificação feita com o modelo logístico binário.

Previsão	Referência	
	Não-PQ	PQ
Não-PQ	180	35
PQ	11	50

Fonte: autora.

Com isso, percebe-se que o modelo mostrou bom desempenho na identificação de não-PQs, já que a vasta maioria dos(as) pesquisadores(as) classificados(as) como não-bolsistas realmente pertence a essa categoria. Além disso, apresentou resultado razoável na classificação de PQs, o que indica que o modelo precisa ser aprimorado especialmente para melhorar a identificação de bolsistas de produtividade. Isso pode ser feito por meio da inclusão de novas variáveis (indicadores de produtividade acadêmica) ou por meio da especificação de métricas já utilizadas, como a diferenciação entre os eventos atendidos no Brasil e no exterior.

6.2 Ajuste do Modelo Logístico Multinomial (Categoria/Nível de PQ)

Considerando as variáveis disponíveis (listadas no Capítulo 4 - *Descrição e Pré-Processamento dos Dados*), foi ajustado um modelo de regressão logística multinomial para modelar o nível e a categoria de produtividade de pesquisadores(as) PQ (isto é, cada um dos grupos 1A, 1B, 1C e 1D, com relação ao grupo 2). As únicas variáveis não incluídas no modelo foram a quantidade total de pós-doutorados (soma de “pós doutorados nacionais” e “pós doutorados internacionais”) e a quantidade total de artigos escritos (soma dos artigos escritos por categoria), sob a justificativa de evitar a multicolinearidade.

Pela variável resposta (nível de produtividade) apresentar relações de ordem entre suas classes, foi considerado o ajuste do modelo ordinal (*Cumulative Logit Models*) (Agresti, 2007). No entanto, o pressuposto de chances proporcionais foi violado - realizou-se o teste de Brant no R, via função *brant* do pacote MASS (R Core Team, 2024) -, logo, a aplicação deste modelo foi desconsiderada.

Inicialmente, empregou-se uma divisão aleatória do conjunto de dados na proporção de 80% para 20%, com o primeiro grupo sendo separado para treinamento do modelo e o segundo, para teste. Então, o ajuste foi realizado com a função *multinom* do pacote *nnet* (Venables e Ripley, 2002) no *software* R. Em seguida, realizou-se um procedimento de retirada das variáveis não-significativas (no modelo) por ordem decrescente do p-valor da seguinte maneira: a variável com maior p-valor associado era removida do conjunto treino (exceto "sexo"), o modelo era reajustado, então a variável não-significativa com maior p-valor associado era removida do conjunto treino, o modelo era novamente reajustado e assim sucessivamente até que só restassem variáveis significativas no modelo. Para essa etapa, foi considerado um nível de significância de 10%. Diferentemente do caso para o modelo logístico binário (Capítulo 6.1), desta vez não temos informações sobre quais variáveis (indicadores de produtividade) são estatisticamente diferentes entre as classes da variável resposta (1D vs. 2, 1C vs. 2, 1B vs 2 e 1A vs. 2), motivo pelo qual a seleção das variáveis se deu de forma diferente neste caso.

Então, ajustou-se um novo modelo considerando as interações de "sexo" com as demais variáveis explicativas, e realizou-se novamente o procedimento de retirada dos termos com coeficientes estatisticamente iguais a zero por ordem decrescente do p-valor (exceto "sexo", por conta das interações), mas considerando um nível de significância de 5% desta vez. Dessa maneira, foi estabelecido o modelo final, contendo as seguintes variáveis: sexo, artigos A1, A3, A4 e B3, orientações para TCC e mestrado em andamento, anos desde o doutoramento e quantidade de pós doutorados feitos no exterior. Além disso, o modelo manteve as interações de sexo com as seguintes variáveis: artigos A3 e A4, TCC, mestrado em andamento e pós doutorados feitos no exterior.

6.2.1 *Classificação e Análise de Produtividade (Categoria/Nível de PQ)*

O intuito da modelagem feita por regressão logística multinomial foi verificar se havia variáveis - dentre as utilizadas - que distinguiam bem docentes PQ de diferentes categorias. Além disso, teve-se por interesse fazer as interpretações para as estimativas dos parâmetros do modelo - em especial, para os coeficientes das interações com o indicador de sexo.

A Tabela 12 apresenta os resultados dos testes de *deviance* do tipo II, obtidos a partir do modelo de regressão logística multinomial ajustado. Esses testes foram aplicados com a função Anova do pacote *car* no software R (Fox e Weisberg, 2019). O procedimento avalia, para cada preditor incluído no modelo, a significância estatística global de sua contribuição para explicar a variável resposta. Para isso, compara-se a qualidade de ajuste do modelo completo com a de um modelo reduzido no qual o preditor em questão é removido. A diferença entre as *deviances* desses modelos é utilizada para calcular a estatística de teste de razão de verossimilhança, cujo valor é então convertido em um p-valor com base na distribuição qui-quadrado, considerando o número de graus de liberdade associado à variável testada. Assim, p-valores baixos (abaixo do nível de significância adotado) indicam que a exclusão do preditor resulta em perda significativa de ajuste, evidenciando que a variável exerce efeito relevante sobre o nível de produtividade científica dos(as) bolsistas analisados(as).

Tabela 12: Variáveis explicativas do modelo logístico multinomial ajustado e p-valores associados (testes de *deviance* do tipo II).

Variável	P-Valor
Sexo	0,0629
A1	0,0000 *
A3	0,0012 *
A4	0,0004 *
B3	0,0015 *
TCC	0,0232 *
Me_and	0,0137 *
posdoc_int	0,0323 *
anos_desde_doc	0,0000 *
A3_Sexo	0,0028 *
A4_Sexo	0,0003 *
TCC_Sexo	0,0000 *
Me_and_Sexo	0,0235 *
posdoc_int_Sexo	0,0001 *

Fonte: autora.

Além disso, ainda considerando o ajuste realizado, os coeficientes e respectivos p-valores por indicador de produtividade e categoria PQ podem ser conferidos na Tabela 13. A distribuição dos(as) bolsistas por categoria revela sub-representação feminina nas categorias mais prestigiadas (apenas 10% em 1A e 15% em 1B, conforme exposto na Tabela 8), o que compromete a equidade e gera instabilidade na modelagem estatística. Coeficientes de alta magnitude para as interações com sexo refletem essa escassez, mas apontam padrões diferenciados de valorização.

Dessa forma, podemos interpretar os coeficientes das variáveis ajustadas (aqueles que se mostraram significativos a 5%). Lembrando que a comparação é estabelecida dentro de cada uma das categorias 1A, 1B, 1C e 1D em relação à categoria de referência - no caso, a categoria 2 de produtividade.

Tabela 13: Coeficientes e p-valores por indicador (variável) e categoria de produtividade

Variável	1A	1B	1C	1D
(Intercept)	-12,9740 (0,0000)*	-8,8013 (0,0000)*	-6,8566 (0,0000)*	-3,2105 (0,0000)*
Sexo	-55,8003 (0,0000)*	-29,6132 (0,0000)*	1,6224 (0,2464)	-1,3596 (0,3374)
A1	0,3040 (0,0000)*	0,1510 (0,0013)*	0,1089 (0,0069)*	0,0137 (0,6957)
A3	0,1950 (0,0096)*	0,0830 (0,1565)	0,2012 (0,0001)*	0,0632 (0,1156)
A4	-0,4030 (0,0215)*	0,1598 (0,1032)	-0,1615 (0,0872)	-0,0133 (0,8366)
B3	0,7720 (0,0511)	-0,8237 (0,1089)	0,3793 (0,2363)	0,0895 (0,7438)
Me_and	-0,5103 (0,3240)	0,1829 (0,5576)	0,5452 (0,0392)*	-0,2809 (0,1923)
TCC	-0,2849 (0,0040)*	-0,1739 (0,0477)*	-0,1332 (0,0805)	-0,1217 (0,0384)*
posdoc_int	0,4828 (0,0697)	0,7334 (0,0025)*	0,5103 (0,0381)*	0,3360 (0,0727)
anos_desde_doc	0,3808 (0,0000)*	0,2490 (0,0000)*	0,1922 (0,0000)*	0,1707 (0,0000)*
A3_Sexo	-68,2060 (0,0000)*	-10,9492 (0,0000)*	-0,1391 (0,6248)	-0,2848 (0,3008)
A4_Sexo	-35,8917 (0,0000)*	11,6095 (0,0000)*	-0,0852 (0,8155)	-0,5051 (0,1998)
posdoc_int_Sexo	77,2864 (0,0000)*	3,7957 (0,0250)*	-0,8746 (0,3973)	1,9834 (0,0476)*
Me_and_Sexo	-85,4648 (0,0000)*	49,5130 (0,0000)*	-0,8176 (0,1824)	-0,3158 (0,5921)
TCC_Sexo	2,8908 (0,0000)*	-86,5079 (0,0000)*	0,1481 (0,2576)	0,1402 (0,2562)

Fonte: autora.

Categoria 1A

- **Sexo:** Mulheres apresentam chance **drasticamente reduzida** de estar na categoria 1A em relação aos homens.
- **Artigos A1:** Para cada publicação, a chance de pertencer à categoria 1A **incrementa em 36%**.
- **Artigos A3:** Para cada publicação, a chance de pertencer à categoria 1A **incrementa em 22%** para os homens e **reduz drasticamente** para as mulheres.
- **Artigos A4:** Para cada publicação, a chance de pertencer à categoria 1A **reduz em 33%** para os homens e **reduz drasticamente** para as mulheres.

- **TCC:** Para cada orientação, a chance de pertencer à categoria 1A **reduz em 25%** para os homens e **incrementa em 14 vezes** para as mulheres.
- **Anos desde doutoramento:** Para cada ano adicional, a chance de pertencer à categoria 1A **incrementa em 46%**.
- **Pós doutorado internacional:** Para cada pós doutorado feito no exterior, a chance das mulheres estarem na categoria 1A **aumenta exponencialmente**.
- **Mestrado em andamento:** Para cada orientação, a chance das mulheres serem classificadas em 1A **reduz drasticamente**.

Categoria 1B

- **Sexo:** Mulheres apresentam chance **drasticamente reduzida** de estar na categoria 1B em relação aos homens.
- **Artigos A1:** Para cada publicação, a chance de pertencer à categoria 1B **incrementa em 16%**.
- **Artigos A3:** Para cada publicação, a chance de pertencer à categoria 1B é **drasticamente reduzida** para as mulheres.
- **Artigos A4:** Para cada publicação, a chance de pertencer à categoria 1B é **consideravelmente incrementada** para as mulheres.
- **TCC:** Para cada orientação, a chance de pertencer à categoria 1B **reduz em 16%** para os homens e **reduz de forma drástica** para as mulheres.
- **Mestrado em andamento:** Para cada orientação, a chance de pertencer à categoria 1B é **exponencialmente incrementada** para as mulheres.
- **Pós doutorado internacional:** Para cada pós doutorado feito no exterior, a chance de pertencer à categoria 1B **incrementa em 2 vezes** para os homens e em **93 vezes** para as mulheres.
- **Anos desde o doutoramento:** Para cada ano adicional, a chance de pertencer à categoria 1B **incrementa em 28%**.

Categoria 1C

- **Artigos A1:** Para cada publicação, a chance de pertencer à categoria 1C **incrementa em 12%**.
- **Artigos A3:** Para cada publicação, a chance de pertencer à categoria 1C **incrementa em 22%**.
- **Mestrado em andamento:** Para cada orientação, a chance de pertencer à categoria 1C **incrementa em 72%**.
- **Pós doutorados internacionais:** Para cada pós doutorado feito no exterior, a chance de pertencer à categoria 1C **incrementa em 67%**.
- **Anos desde o doutoramento:** Para cada ano adicional, a chance de pertencer à categoria 1C **incrementa em 21%**.

Categoria 1D

- **TCC:** Para cada publicação, a chance de pertencer à categoria 1D **reduz em 11%**.
- **Ano desde doutoramento:** Para cada ano adicional, a chance de pertencer à categoria 1D **incrementa em 19%**.
- **Pós doutorados internacionais:** Para cada pós doutorado feito no exterior, as chances de classificação em 1D **incrementam em 7 vezes** para as mulheres.

A partir das interpretações expostas, é possível notar alguns padrões. Por exemplo: para cada ano adicional desde o doutoramento, a chance de pertencer a cada uma das categorias 1D, 1C, 1B e 1A, em relação à categoria 2, aumenta com o prestígio da categoria (incrementos de 19%, 21%, 28% e 46%, respectivamente), apresentando o menor valor para 1D (19%) e o maior, para 1A (46%). Isso evidencia a importância do tempo de experiência para o recebimento da bolsa e para a valorização da produção do(a) pesquisador(a), produção essa que provavelmente se acumula e aperfeiçoa com os anos.

Ademais, o mesmo efeito de incremento da chance com o prestígio da classe PQ pode ser notado para a publicação de artigos A1 nas categorias 1C, 1B e 1A - para cada publicação, as chances de pertencimento a essas categorias incrementam em 12%, 16% e 36%, respectivamente. Além disso, para cada publicação de artigo A3 (uma produção característica de homens PQ, conforme evidenciado na Tabela 7), a chance de pertencimento à classe mais prestigiada (1A) aumenta em 22% para os homens enquanto reduz exponencialmente para as mulheres. Esses resultados demonstram a força de publicações de alto nível (A1 e A3) na classificação de pesquisadores(as) em níveis de produtividade mais prestigiados, e como os homens são mais valorizados por suas produções.

Os coeficientes estimados também evidenciam que a publicação de artigos em periódicos A4 não é característica de PQs da classe 1A, já que, para cada publicação, a chance de pertencer à 1A (em relação à classe 2) reduz em 33% para os homens e de forma exponencial para as mulheres.

Adicionalmente, quando consideramos o indicador "TCC", percebemos que o aumento do número de orientações está relacionado a um aumento das chances de classificação no nível 2, o nível de produtividade menos prestigiado. A exceção ocorre na categoria 1A de produtividade, para a qual as mulheres parecem ser mais valorizadas por esse tipo de orientação - no entanto, os resultados opostos obtidos para as demais categorias (coeficientes negativos para 1B e 1D) podem sugerir que talvez as orientações para TCC só sejam fortemente presente entre as mulheres bolsistas da categoria 1A, o que acaba sendo refletido pelo ajuste do modelo.

Considerando o grupo de pesquisadores(as) bolsistas, é importante notar que o TCC é uma orientação "feminina" (Tabela 7) e que, no geral, é entendida pelo modelo como característica da categoria menos prestigiada de produtividade (categoria 2). Enquanto isso, a publicação de artigos A3 (produção "masculina") está relacionada apenas a incrementos da chance de classificação na categoria de produtividade mais prestigiada (1A) para os homens. Talvez esses resultados sejam reflexo de vieses de gênero presentes no processo de avaliação e distribuição das bolsas.

Os resultados do modelo multinomial também indicam que o pós-doutorado internacional é um fator fortemente associado à classificação em diferentes categorias da bolsa de produtividade do CNPq, e apresenta pesos diferentes entre homens e mulhe-

res: entre as mulheres, cada experiência de pós-doutorado no exterior está associada a um aumento exponencial nas chances de alcançar a categoria 1A, a um incremento de 93 vezes nas chances de atingir a 1B (em contraste com apenas 2 vezes para os homens), a um aumento de 67% na 1C (para ambos os sexos) e a um aumento de 7 vezes nas chances de enquadramento na 1D. Esses achados sugerem que, embora o pós-doutorado internacional seja um diferencial valorizado, sua influência não garante necessariamente a inserção nas categorias mais altas para todas as pesquisadoras. Isso pode ser injusto para com as mulheres, em razão de fenômenos como a dupla jornada de trabalho e a sobrecarga com responsabilidades domésticas e familiares, que dificultam a conversão desse capital acadêmico em progressão para níveis mais prestigiados, mesmo possuindo qualificações equivalentes às de seus pares masculinos.

De forma similar ao método empregado no Capítulo 6.1.1 - *Classificação e Análise de Produtividade (PQs e Não-PQs)*, foi realizada a análise da matriz de correlações para as variáveis explicativas do modelo - além da análise dos valores calculados para o VIF - com o intuito de verificar se havia evidências de multicolinearidade. No entanto, as estatísticas calculadas sugeriram o contrário (correlações abaixo de 0,8 e valores de VIF abaixo de 10).

Por fim, realizou-se a classificação para os(as) pesquisadores(as) da base-teste. Os resultados podem ser conferidos na Tabela 14 e a acurácia geral do modelo foi de 55% (IC 95%: (0.44,0.66)). As demais métricas da classificação são mostradas na Tabela 15. De forma geral, os resultados para a classificação se mostraram insatisfatórios. Uma sugestão para trabalhos futuros seria considerar técnicas como o balanceamento das classes da variável resposta (He e Ma, 2013), já que a distribuição de PQs nas categorias de produtividade é desigual, com a maioria se concentrando nos níveis 1D e 2 (Tabela 8).

Tabela 14: Resultados da classificação feita com o modelo logístico multinomial.

Previsão	Referência				
	1A	1B	1C	1D	2
1A	1	0	3	2	0
1B	0	2	0	1	0
1C	1	0	0	3	4
1D	3	3	3	13	3
2	0	1	1	9	30

Fonte: autora.

Tabela 15: Métricas por classe para o modelo logístico multinomial.

Classe	Sensibilidade	Especificidade	Acurácia Balanceada	F1 Score
1A	0.2000	0.9359	0.5679	0.182
1B	0.3333	0.9870	0.6601	0.444
1C	0.0000	0.8947	0.4474	–
1D	0.4643	0.7818	0.6231	0.490
2	0.8108	0.7609	0.7858	0.769

Fonte: autora.

Pelas métricas calculadas (Tabela 15), percebe-se que a classe 2 foi aquela para a qual o modelo apresentou melhor desempenho (acurácia balanceada de 79%), seguida pelas classes 1B (66%) e 1D (62%). Já para as classes 1A e 1C, o modelo apresentou baixo desempenho (acurácias balanceadas de 57% e 45%). Com relação à *f1 score*, estatística que combina valor preditivo positivo (precisão) e sensibilidade (Bramer, 2016), essa se mostrou mais alta para a classe 2 (77%) e muito baixa para as demais classes - em especial, para a classe 1A (18%).

Com isso, percebe-se que o modelo apresenta o melhor desempenho na identificação das classes menos prestigiadas (1D e 2) e da classe 1B, enquanto desempenha de forma insatisfatória na identificação da classe mais prestigiada (1A) e da classe 1C.

7 DISCUSSÃO DOS RESULTADOS

Neste estudo, teve-se por objetivo a busca por critérios objetivos que distinguíssem pesquisadores(as) bolsistas de produtividade de pesquisadores(as) não-bolsistas. Além disso, considerando apenas os(as) pesquisadores(as) PQ, buscou-se também identificar critérios de distinção entre os(as) bolsistas das categorias 1A, 1B, 1C, 1D e 2. Por fim, com base nesses critérios, intencionou-se verificar se era possível classificar pesquisadores(as) em bolsistas ou não-bolsistas - além de classificar a categoria de produtividade no caso dos(as) pesquisadores(as) bolsistas -, e fazer a interpretação dos coeficientes do modelo em termos de produtividade e sexo, a fim de verificar se havia evidências de vieses de gênero embutidos no ajuste.

7.1 Perfis Científicos e Desigualdade de Reconhecimento entre Sexos

Com relação às análises inferenciais e preditivas realizadas para PQs e Não-PQs, obteve-se que os(as) PQs publicam mais artigos, livros e capítulos de livros; orientam mais alunos de graduação e pós-graduação, possuem mais anos de experiência (tempo desde o doutoramento), mais pós-doutorados feitos no exterior e maior participação em eventos acadêmicos. Em suma: PQs apresentam desempenho mediano melhor para 18 dos 25 indicadores considerados. Esses achados refletem que, de fato, existem aspectos objetivos (métricas de produção acadêmica) que distinguem pesquisadores(as) receptores da bolsa de produtividade em pesquisa dos(as) demais pesquisadores(as).

Considerando agora os testes aplicados para investigar diferenças de produção entre sexos, constatou-se que homens e mulheres apresentam diferentes perfis científicos: homens publicam mais artigos de alto prestígio (A3), enquanto mulheres orientam mais para graduação (IC e TCC) e publicam mais capítulos de livros. Além disso, na etapa da modelagem, constatou-se que mulheres também tendem a ser classificadas em níveis menos prestigiados de produtividade pelo simples fato de serem mulheres, apresentando chance muito reduzida de estar nas duas categorias de maior prestígio (1A e 1B) em relação aos homens.

Ademais, considerando os resultados obtidos para homens e mulheres do grupo de PQs, as evidências apontam que as mulheres precisam produzir mais para obter

o mesmo reconhecimento que seus pares masculinos, já que apresentaram desempenho superior para 4 dos 5 indicadores de produtividade que se mostraram estatisticamente diferentes entre sexos, mesmo sendo a minoria dos(as) pesquisadores(as) bolsistas de produtividade (12% são mulheres e 88% são homens).

7.2 Sobre as Classificações (Modelagens)

Neste trabalho, pelo foco da modelagem não ser apenas o preditivo, mas também o inferencial, priorizou-se técnicas da Estatística Clássica no ajuste dos modelos logístico binário e multinomial. No entanto, se o foco fosse o classificatório, poderiam ter sido adotadas abordagens de Aprendizado de Máquina, como a validação cruzada e a floresta aleatória (Izbicki e Santos, 2020), além do balanceamento das classes da variável resposta (He e Ma, 2013).

Com relação ao modelo logístico binário simples, utilizado para distinguir pesquisadores(as) PQ dos(as) Não-PQ, obteve-se valores razoavelmente bons para a acurácia balanceada (76,5%) e o *f1 score* (68%). Na verdade, o que foi constatado é que o modelo apresenta excelente capacidade de identificação de pesquisadores(as) não-bolsistas (especificidade de 94%), mas enfrenta um pouco de dificuldade na identificação de bolsistas (sensibilidade de 59%). Além disso, considerando a classificação feita para a categoria/nível de produtividade de pesquisadores(as) bolsistas (modelagem logística multinomial), observou-se que o modelo apresentou bom desempenho na identificação da classe menos prestigiada - isto é, a classe 2 (acurácia balanceada de 79%)-, desempenho razoável na identificação das classes 1B (66%) e 1D (62%), além de desempenho insatisfatório na classificação das classes 1A (57%) e 1C (45%).

Esses resultados possivelmente se justificam, em parte, pelo fato de que os indicadores avaliados neste estudo contemplam apenas 2 dos 5 critérios de julgamento para seleção dos(as) bolsistas de produtividade (são eles: critérios 1 e 3. Verificar na Tabela 1). Aspectos como a qualidade do projeto de pesquisa, a participação do(a) pesquisador(a) em sociedades e grupos de pesquisa no país e no exterior, as contribuições na editoria de periódicos e as condecorações também são critérios levados em consideração no julgamento mas aos quais não tivemos acesso. Além disso, em 2024, o CNPq estabeleceu novas categorias de produtividade (A, B e C), que representam agrupamentos das categorias antigas (a nova classe A abrange os antigos níveis 1A

e 1B, enquanto a classe B abrange 1C e 1D, e a classe C engloba o antigo nível 2) (Brasil, 2024). Talvez essa reformulação se deva ao fato das categorias antigas não estarem distinguindo bem pesquisadores(as).

7.3 Considerações Finais

Neste trabalho, um dos achados mais relevantes foi a constatação da baixa representatividade feminina nos Programas de Pós-Graduação (PPGs) de Matemática, Estatística e Probabilidade, nos quais apenas 20% dos(as) pesquisadores(as) são mulheres. Essa sub-representação pode refletir tanto barreiras históricas de acesso, permanência e progressão dessas pesquisadoras na área quanto aspectos culturais relacionados à forma como o gênero é socialmente construído, o que impacta diretamente o incentivo e a presença feminina nas áreas de STEM. Os impactos dessa desigualdade extrapolam o âmbito individual, pois a ausência de diversidade limita perspectivas de pesquisa, restringe a inovação e perpetua assimetrias no reconhecimento científico.

Os resultados obtidos neste estudo também evidenciam que, embora existam métricas objetivas claras associadas ao recebimento da bolsa de produtividade, tais métricas tendem a valorizar mais as produções características dos homens, enquanto certas produções mais frequentes entre as mulheres chegam a ser desvalorizadas na modelagem - o que indica a presença de desigualdades estruturais no reconhecimento científico.

A relevância deste estudo reside justamente em trazer à luz essas assimetrias, permitindo um debate mais fundamentado sobre as formas de avaliação da produtividade acadêmica e seus potenciais vieses. Ao combinar análise quantitativa rigorosa com a discussão sobre gênero na ciência, este trabalho contribui para ampliar a compreensão de como critérios aparentemente objetivos podem reproduzir desigualdades históricas, servindo como subsídio para políticas públicas e institucionais mais equitativas.

No entanto, algumas limitações precisam ser reconhecidas. A ausência da informação da raça dos(as) pesquisadores(as) impediu a análise específica sobre mulheres negras, grupo que historicamente enfrenta barreiras ainda mais acentuadas na ciência. Além disso, o indicador “eventos” não permitiu distinguir participações nacionais de internacionais, o que poderia fornecer nuances relevantes para a interpretação dos

resultados.

Apesar dessas restrições, o estudo oferece contribuições significativas tanto para a comunidade acadêmica quanto para a sociedade. Para a ciência, fornece evidências empíricas que podem embasar discussões sobre critérios de avaliação mais inclusivos e representativos. Para a sociedade, reforça a importância da equidade de gênero na produção e reconhecimento científicos, apontando caminhos para reduzir disparidades e ampliar a diversidade no cenário acadêmico brasileiro.

REFERÊNCIAS

AGRESTI, A. **An Introduction to Categorical Data Analysis**. 2 ed. New York: Wiley, 2007.

BEAUVOIR, Simone de. **Le deuxième sexe**, Vol. 2. Paris: Gallimard, 1949.

BOTTALLO, Ana. **Mulheres são apenas 36% de bolsistas de produtividade do CNPq há 20 anos**. São Paulo: Folha de S. Paulo, 03 jan. 2024. Disponível em: <https://www1.folha.uol.com.br/educacao/2024/01/mulheres-sao-apenas-36-de-bolsistas-de-produtividade-do-cnpq-ha-20-anos.shtml>. Acesso em: 16 jun. 2025.

BRAMER, M. **Principles of Data Mining**. 3 ed. Londres: Springer-Verlag London Ltd, 2016.

BRASIL. **Bolsas de Produtividade: CNPq investe 50% mais em relação ao edital do ano passado; submissões até 30/12**. Resolução Normativa 12/2024. Brasília: CNPq, 2024. Disponível em: <https://www.gov.br/cnpq/pt-br/assuntos/noticias/cnpq-em-acao/cnpq-anuncia-chamada-para-concessao-de-bolsas-de-produtividade>. Acesso em: 11 ago. 2025.

BRASIL. **Professoras são minoria em seis grandes áreas do conhecimento**. Publicada em: 01 abr. 2025. Brasília: CAPES, 2025. Disponível em: <https://www.gov.br/capes/pt-br/assuntos/noticias/professoras-sao-minoria-em-seis-grandes-areas-do-conhecimento>. Acesso em: 16 jun. 2025.

BUSSAB, W.O.; MORETTIN, P.A. **Estatística Básica**. 6 ed. São Paulo: Saraiva, 2010.

CNN. **CNN Innovation: o que há por trás dos erros da Inteligência Artificial?**. Publicada em: 23 mar. 2025. Disponível em: <https://www.cnnbrasil.com.br/tecnologia/cnn-innovation-o-que-ha-por-tras-dos-erros-da-inteligencia-artificial/>. Acesso em: 17 jun 2025.

CONOVER, W.J. **Practical Nonparametric Statistics**. 3 ed. New York: John Wiley & Sons, Inc., 1999.

COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR (CAPES). **Documento Técnico do QUALIS Periódicos**. Disponível em: <https://www.gov.br/capes/pt-br/centrais-de-conteudo/documentos/avaliacao/avaliacao-quadrienal-2017/DocumentotecnicoQualisPeridicosfinal.pdf>. Acesso em 03 jul. 2025.

COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR (CAPES). **RELATÓRIO DE AVALIAÇÃO 2017-2020 (QUADRIENAL 2021)**.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO (CNPq). **Chamada CNPq nº 09/2023: Bolsas de Produtividade em Pesquisa – PQ**. Brasília, 2023. Disponível em: http://memoria2.cnpq.br/web/guest/chamadas-publicas?p_p_id=resultadosportlet_WAR_resultadoscnpqportlet_INSTANCE_0ZaM&filtro=abertas&detalha=chamadaDivulgada&idDivulgacao=11445. Acesso em: 16 jun. 2025.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO (CNPq). Página institucional. Brasília, 2025. Disponível em: <https://www.gov.br/cnpq/pt-br>. Acesso em: 16 jun. 2025.

CORALINA, Cora. **Carta enviada à filha**. Brasil: Acervo pessoal, 1980.

DE MARI, D. **Elas não querem ser cientistas**. São Paulo, 29 jan. 2025. Disponível em: <https://minabemestar.uol.com.br/meninas-nas-areas-stem/>. Acesso em: 16 jun. 2025.

EVERITT, B. S. **The analysis of contingency tables**. 2 ed. London: Chapman & Hall, 1992.

FIELD, A. P.; MILES, J., & FIELD, Z. **Discovering statistics using R**. London: Andy Field, Jeremy Miles and Zoë Field, 2012.

FORÇA MENINAS. **Meninas curiosas, mulheres de futuro**. 2023. Disponível em: <https://frmeninas.com.br/pesquisa-meninas-brasileiras-e-a-insercao-em-stem/>. Acesso em: 16 jun. 2025.

FOX, J.; WEISBERG, S. **An R Companion to Applied Regression**. 3 ed. Thousand Oaks (CA): Sage, 2019.

FREIRE, T. **Ausência de Mulheres Negras é Desafio para a Ciência**. Rio de Janeiro: Agência Brasil, 11 fev. 2025. Disponível em: <https://agenciabrasil.ebc.com.br/direitos-humanos/noticia/2025-02/ausencia-de-mulheres-negras-e-desafio-para-ciencia>. Acesso em 10 ago. 2025.

GALEANO, Eduardo. **Las palabras andantes**. 5. ed. Buenos Aires: Catálogos S.R.L., 2001.

GIBBONS, J.D., CHAKRABORTI, S. **Nonparametric Statistical Inference**. 4 ed. New York: Marcel Dekker, 2003.

HE, H.; MA, Y. **Imbalanced Learning: Foundations, Algorithms, and Applications**. 1 ed. New York: The Institute of Electrical and Electronics Engineers, Inc., 2013.

HOSMER, D.W.; LEMESHOW, S; STURDIVANT, R.X. **Applied Logistic Regression**. 3 ed. Hoboken, New Jersey: John Wiley & Sons, Inc., 2013.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Censo demográfico 2022: resultados**. Rio de Janeiro, 2022. Disponível em: <https://censo2022.ibge.gov.br/>. Acesso em: 10 ago. 2025.

IBAÑES, M. V. **Acompanhamento e avaliação em alguns instrumentos de fomento à pesquisa do CNPq**. Brasília: CNPq/DCT, 1996.

IZBICKI, R.; DOS SANTOS, T.M. **Aprendizado de máquina: uma abordagem estatística**. 1 ed. São Carlos: Câmara Brasileira do Livro, 2020.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An introduction to statistical learning: with applications in R**. 2. ed. New York: Springer, 2021. Disponível em: <https://www.statlearning.com>. Acesso em: 28 jul. 2025.

LOPES, C. de M. N.; INFANGER, C.; STARLINO, J. A. B.; MARQUES, K. M.; CURY, A. A.; SARMANHO, A. M. C.; MENDES, J. C. **A representatividade feminina nos programas brasileiros de pós-graduação em engenharia civil, ambiental e de transportes**. Caderno Pedagógico, v. 21, n. 3, 2024.

MCKINSEY & COMPANY. **Women in the workplace**. São Paulo, 2020. Disponível em: <https://www.mckinsey.com/featured-insights/diversity-and-inclusion/women-in-the-workplace>. Acesso em: 16 jun. 2025.

MONTGOMERY, D.C., PECK, E.A., VINING, G.G. **Introduction to Linear Regression Analysis**. 5 ed. Hoboken, New Jersey: John Wiley & Sons, Inc., 2012.

NERY, I.F. **Por que tão poucas? Análise do Efeito Tesoura e Leaky Pipeline na carreira de mulheres pesquisadoras na engenharia**. Tese (Mestrado em Engenharia Civil) - Faculdade de Engenharia Civil, Universidade Federal de Juiz de Fora, Juiz de Fora, 2025.

PUCPR DIGITAL. **Como a dupla jornada de trabalho afeta a saúde mental da mulher?**. Publicada em: 8 mar. 2023. Disponível em: <https://posdigital.pucpr.br/blog/dupla-jornada-de-trabalho>. Acesso em 14 de jul. 2025.

QUEIROZ, D. **Bacharelado em Matemática é mais Desigual para Mulheres Negras**. São Paulo: Jornal da USP, 16 jun. 2023. Disponível em: <https://jornal.usp.br/diversidade/bacharelado-em-matematica-e-mais-desigual-para-mulheres-negras/>. Acesso em: 10 ago. 2025.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria: R Foundation for Statistical Computing, 2024. Disponível em <https://www.R-project.org/>. Acesso em 28 jul. 2025.

SANTAMARÍA, L.; MIHALJEVIĆ, H. **Comparison and benchmark of name-to-gender inference service**. PeerJ Comput Sci, v. 16, n. 4, p. e156, 2018.

SILVA, L.L. **Estudo do Perfil Científico dos Pesquisadores com Bolsa de Produtividade do CNPq que atuam no Ensino de Ciências e Matemática**. Revista Brasileira de Pesquisa em Educação em Ciências, v.11, n.3, p. 75-99, 2011.

VENABLES, W. N.; RIPLEY, B.D. **Modern Applied Statistics with S**. 4 ed. New York: Springer, 2002.

APÊNDICE A - PPGs Contemplados

Tabela A.1: Programas de Pós-Graduação em Matemática, Estatística e Probabilidade contemplados no estudo.

Nome do Programa	Sigla IES	Nível
Modelagem Matemática	FGV/RJ	ME/DO
MATEMÁTICA	FUFPI	ME/DO
MATEMÁTICA	IMPA	ME/DO
MATEMÁTICA APLICADA E COMPUTACIONAL	PUC-RIO	ME/DO
MATEMÁTICA	UEL	ME
MATEMÁTICA	UEM	ME/DO
MATEMÁTICA	UFABC	ME/DO
MATEMÁTICA	UFAL	ME
MATEMÁTICA	UFAM	ME
MATEMÁTICA - UFBA - UFAL	UFBA	ME
MATEMÁTICA	UFBA	DO
MATEMÁTICA - UFPB - J.P. - UFCG	UFC	ME/DO
MATEMÁTICA	UFCG	DO
MATEMÁTICA	UFCG	ME
MATEMÁTICA	UFES	ME
MATEMÁTICA	UFF	ME/DO
MATEMÁTICA	UFG	ME/DO
MATEMÁTICA	UFJF	ME
MATEMÁTICA	UFMA	ME
ESTATÍSTICA	UFMG	ME/DO
MATEMÁTICA E ESTATÍSTICA	UFMG	ME/DO
MATEMÁTICA - UFPA - UFAM	UFPA	ME
MATEMÁTICA	UFPA	DO
MATEMÁTICA	UFPB-JP	ME/DO
ESTATÍSTICA	UFPE	ME/DO

Continuação da Tabela A.1.

Nome do Programa	Sigla IES	Nível
MATEMÁTICA	UFPE	ME/DO
MATEMÁTICA APLICADA	UFRGS	ME/DO
ESTATÍSTICA	UFRGS	ME/DO
MATEMÁTICA	UFRGS	ME
ESTATÍSTICA	UFRJ	ME/DO
MATEMÁTICA APLICADA E ESTATÍSTICA	UFRJ	ME/DO
MATEMÁTICA PURA E APLICADA	UFRN	ME
MATEMÁTICA	UFSC	ME/DO
ESTATÍSTICA	UFSCAR	ME/DO
MATEMÁTICA	UFSCAR	ME/DO
MATEMÁTICA	UFSM	ME
MATEMÁTICA	UFU	ME
MATEMÁTICA	UFV	ME
ESTATÍSTICA	UNB	ME/DO
MATEMÁTICA APLICADA E COMPUTACIONAL	UNB	ME
MATEMÁTICA	UNESP-PP	ME
MATEMÁTICA	UNESP-SJRP	ME/DO
MATEMÁTICA APLICADA	UNICAMP	ME/DO
ESTATÍSTICA	UNICAMP	ME/DO
MATEMÁTICA	UNICAMP	ME/DO
MATEMÁTICA PURA E APLICADA	UNIFEI	ME
MATEMÁTICA	UNIFESP	ME
MATEMÁTICA APLICADA	USP	ME/DO
ESTATÍSTICA	USP	ME/DO
MATEMÁTICA	USP	ME/DO
MATEMÁTICA	USP/RP	ME

Continuação da Tabela A.1.

Nome do Programa	Sigla IES	Nível
MÉTODOS MATEMÁTICOS EM FINANÇAS	USP/SC	ME/DO
MATEMÁTICA	IMPA	MP
MATEMÁTICA APLICADA E COMPUTACIONAL	UNESP-RC	MP
Ensino de Matemática	UNICAMP	MP
Matemática, Estatística e Computação Aplicadas à Indústria	USP/SC	MP

Fonte: autora (2025).

APÊNDICE B - PQs e Não-PQs

Tabela B.1: Indicadores estatisticamente (5%) iguais entre PQs e Não-PQs.

Figura	Indicador	Valor-p (mediana)
2	Artigo B4	0.5125
2	Artigo C	0.3980
4	Cong	0.2874
7	IC and	0.6822
7	Mestrado and	0.1020
7	TCC	0.0782
9	Pós Docs BR	1

Fonte: autora (2025).

Tabela B.2: Estatísticas para os indicadores de PQs e Não-PQs. Indicadores com asterisco (*) se mostraram significativos no teste da mediana.

Indicador	Mediana (PQs)	Mediana (Não-PQs)	Mediana conjunta	PQs (%) acima da mediana	Não-PQs (%) acima da mediana
Artigo A1*	4	1	1	78	29
Artigo A2*	6	1	2	80	29
Artigo A3*	4	1	1	75	38
Artigo A4*	2	1	1	56	35
Artigo B1*	0	0	0	44	37
Artigo B2*	0	0	0	35	26
Artigo B3*	0	0	0	17	13
Artigo B4	0	0	0	13	12
Artigo C	0	0	0	22	20
Total Artigos*	20	7	9	91	31
Cong	2	2	2	49	46
Livros*	0	0	0	23	13
Cap. Livros*	0	0	0	36	24

Indicador	Mediana (PQs)	Mediana (Não- PQs)	Mediana conjunta	PQs (%) acima da mediana	Não-PQs (%) acima da mediana
IC and	0	0	0	23	24
IC conc*	4	2	3	50	42
Me and	0	0	0	39	35
Mestrado conc*	5	2	3	61	38
Doc and*	1	0	0	74	37
Doutorado*	4	0	1	72	29
TCC	0	0	0	34	39
Anos Doc*	16	14	15	54	44
Total Pós Docs*	1	1	1	41	32
Pós Docs Exterior*	1	0	0	62	43
Pós Docs BR	0	0	0	44	44
Eventos*	13	8	9	63	42

Continuação da Tabela B.2

Fonte: autora (2025).

APÊNDICE C - Mulheres e Homens

Tabela C.1: Indicadores estatisticamente (5%) iguais entre homens e mulheres (caso geral).

Figura	Indicador	Valor-p (mediana)
15	Artigos B1	0.5641
15	Artigos B2	0.9856
15	Artigos B3	0.2267
15	Artigos B4	0.08008
15	Artigos C	0.7446
20	Mestrado conc	0.4183
17	Cong	0.8761
18	Livros	0.5883
21	Anos desde doc	0.8761
22	Pós Docs Int	0.8989
22	Pós Docs BR	0.9633
22	Pós Docs Total	0.3012

Fonte: autora (2025).

Tabela C.2: Estatísticas para os indicadores de mulheres (F) e homens (H). Indicadores com asterisco (*) se mostraram significativos no teste da mediana.

Indicador	Mediana (F)	Mediana (H)	Mediana conjunta	F (%) acima da mediana	H (%) acima da mediana
Artigo A1*	1	1	1	35	46
Artigo A2*	1	2	2	33	48
Artigo A3*	1	2	1	39	52
Artigo A4*	1	1	1	30	44
Artigo B1	0	0	0	38	39
Artigo B2	0	0	0	29	29

Indicador	Mediana (F)	Mediana (H)	Mediana conjunta	F (%) acima da mediana	H (%) acima da mediana
Artigo B3	0	0	0	17	14
Artigo B4	0	0	0	16	12
Artigo C	0	0	0	22	21
Total Artigos*	7	10	9	37	53
Cong	2	2	2	48	47
Livros	0	0	0	18	16
Cap. Livros*	0	0	0	39	25
IC and*	0	0	0	30	22
IC*	5	2	3	59	41
Me and*	0	0	0	43	35
Mestrado conc	3	3	3	47	44
Doc and*	0	0	0	42	49
Doutorado conc*	1	1	1	34	45
TCC*	0	0	0	49	35
Anos Doc	15	15	15	48	47
Total Pós Docs	1	1	1	32	35
Pós Docs Exterior	0	0	0	44	49
Pós Docs BR	0	0	0	49	44
Eventos*	9	12	9	55	47

Continuação da Tabela C.2

Fonte: autora (2025).

Tabela C.3: Estatísticas para os indicadores significativos entre mulheres (F) e homens (H) bolsistas de produtividade (PQs).

Indicador	Mediana (F)	Mediana (H)	Mediana conjunta	F (%) acima da mediana	H (%) acima da mediana
Artigo A3*	3	4	4	23	41
Doutorado conc*	4	5	3.5	53	37
IC conc*	4	6	3	58	41
TCC*	0	0	0	49	32
Cap. Livros*	0	1	0	57	33

Fonte: autora (2025).

Tabela C.4: Estatísticas para os indicadores significativos entre mulheres (F) e homens (H) por categoria da bolsa de produtividade.

Grupo PQ	Indicador	Mediana (F)	Mediana (H)	Mediana conjunta	F (%) acima da mediana	H (%) acima da mediana
1D	Pós Doc Total*	1	2	1	85	44
1D	Eventos*	45	14	15	77	44
2	Cap Livros*	1	0	0	68	29
2	Doutorado conc*	4	5	2	68	36

Fonte: autora (2025).

APÊNDICE D - Função do Teste da Mediana

```
median.test2sample <- function(X,Y){  
  AmComb=c(X,Y)  
  medAC=median(AmComb)  
  A=sum(X>medAC)  
  B=sum(Y>medAC)  
  m=length(X)  
  n=length(Y)  
  C=m-A  
  D=n-B  
  tabMed=matrix(c(A,C,B,D),2,2)  
  tabMed  
  stab=sum(tabMed)  
  if (stab< 20) test= fisher.test(tabMed)  
  else test=chisq.test(tabMed)  
  return(test)  
}
```