

O Impacto de Bolsas de Doutorado na Pós-Graduação Brasileira: Uma análise pelo Currículo Lattes do CNPq

*The Impact of Doctoral Scholarships on Brazilian Graduate Studies:
An Analysis of the CNPq Lattes Curriculum*

*El impacto de las becas de doctorado en los estudios de posgrado brasileños:
un análisis del currículo de CNPq Lattes*

ALBERTO DE CAMPOS E SILVA¹

ALEXANDRE GUILHERME MOTTA SARMENTO²

Resumo: Este artigo calculou um impacto do financiamento em bolsas de doutorado pela CAPES e pelo CNPq por meio de análise cientométrica com data mining nos Currículos Lattes. Assim, foram retirados 140 distintos indicadores de 21.250 currículos, nos quais se aplicou modelo econométrico quase-experimental, por meio de regressão linear multivariada em cada indicador e o método de diferença em diferenças, como recomendado pelo Banco Mundial, chegando-se em análise custo-efetividade de cada indicador significativo. Encontrou-se impacto de 15.019 artigos publicados e 8.927 trabalhos em eventos, entre outros.

Palavras-chave: Avaliação de Impacto; Cientometria; Bolsas de Doutorado; Currículo Lattes; Indicadores Bibliométricos.

Abstract: *This article calculated the impact of funding on doctoral scholarships from CAPES and CNPq through scientometric analysis with data mining in Lattes Curriculum. Thus, 140 different indicators were taken from 21,250 curriculum vitae, to which a quasi-experimental econometric model was applied, using multivariate linear regression on each indicator and the difference-in-differences method, as recommended by the World Bank, arriving at a cost-effectiveness analysis of each significant indicator. The impact of 15,019 published articles and 8,927 works at events was registered, among other issues.*

Keywords: *Impact Evaluation; Scientometrics; PhD Scholarships; Lattes Curriculum; Bibliometric Indicators.*

1 ORCID: <https://orcid.org/0000-0001-6506-1988> . Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Universidade Federal do Rio Grande do Sul (UFRGS), Programa de Pós-Graduação em Educação em Ciências, Brasília, DF, Brasil.

2 ORCID: <https://orcid.org/0000-0002-7201-4466> . Universidade Federal do Rio Grande do Sul (UFRGS), Programa de Pós-Graduação em Educação em Ciências. Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)., Brasília, DF, Brasil.

Resumen: *Este trabajo analizó el impacto del financiamiento de becas de doctorado otorgadas por la CAPES y el CNPq mediante un análisis cuantitativo con extracción de datos del Currículo Lattes. Así, a partir del análisis de 21.250 currículos, se encontró 140 indicadores diferentes, a los que se aplicó un modelo econométrico cuasiexperimental. Este modelo empleó regresión lineal multivariada para cada indicador y el método de diferencias en diferencias, recomendado por el Banco Mundial, con el fin de realizar un análisis de costo-efectividad para cada indicador significativo. Con el resultado, se observó el impacto de 15.019 artículos publicados y 8.927 trabajos presentados en eventos, entre otros resultados relevantes.*

Palabras clave: *Evaluación de Impacto; Cuantitativa; Becas de Doctorado; Currículo Lattes; Indicadores Bibliométricos.*

INTRODUÇÃO

Ao formar doutores, as universidades e instituições de pesquisa contribuem diretamente na qualidade da educação no país, produzindo profissionais capazes de desenvolver pesquisas relevantes e inovadoras e contribuindo para o desenvolvimento social e econômico do país. Como exemplo, o nível de ensino de professores da educação básica evoluiu juntamente com o crescimento da pós-graduação *stricto sensu*, mostrando a importância e a preocupação com formação continuada de professores. Apesar disso, mesmo com ações de fomento se acumulando ao investimento natural nos últimos anos, o sistema educacional carece de maior esforço em sua cadeia de formação para a qualificação de seus profissionais e o respectivo aumento de professores pós-graduados (Silva; Porciúncula, 2017).

Nesse sentido, o Plano Nacional de Educação (PNE) determina diretrizes, metas e estratégias para a política educacional no período de 2014 a 2024 no intuito de promover a formação de mestres e doutores. Entre as 20 metas do Plano, as de nº 13 e 14 envolvem a formação de doutores no país: “elevar a qualidade da educação superior e ampliar a proporção de mestres e doutores do corpo docente em efetivo exercício no conjunto do sistema de educação superior para 75% (setenta e cinco por cento), sendo, do total, no mínimo, 35% (trinta e cinco por cento) doutores” e “elevar gradualmente o número de matrículas na pós-graduação de modo a atingir a titulação anual de 60.000 (sessenta mil) mestres e 25.000 (vinte e cinco mil) doutores” (PNE - Plano Nacional de Educação, n.d.).

De acordo com o MEC, um dos aspectos que influencia a qualidade da educação superior é a titulação do corpo docente, sobretudo em cursos de mestrado e doutorado. Uma das estratégias para o cumprimento da meta relacionada é elevar o padrão de qualidade das universidades, visando à realização de pesquisa institucionalizada de modo articulado a programas de pós-graduação *stricto sensu*. Entre as estratégias estabelecidas no novo PNE, encontram-se expandir o financiamento da pós-graduação *stricto sensu* por meio das agências oficiais de fomento, incentivar a atuação em rede e o fortalecimento de grupos de pesquisa e

promover o intercâmbio científico e tecnológico, nacional e internacional, entre as instituições de ensino, pesquisa e extensão (Presidência da República; Ministério da Educação; Secretaria de Articulação com os Sistemas de Ensino, 2014).

Assim, é na pós-graduação, principalmente no nível de doutorado, que pesquisa e educação possuem forte área de intersecção. Nesse sentido, as estratégias de fortalecimento da pós-graduação do governo federal do Brasil competem a duas agências de fomento: a CAPES e o CNPq. A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) é uma Fundação do Ministério da Educação (MEC) e tem como missão a expansão e consolidação da pós-graduação *stricto sensu* (mestrado e doutorado) no Brasil (História e Missão – CAPES, [s. d.]). Para cumprir seu objetivo, a CAPES concede bolsas de estudo no Brasil visando à formação de doutores. Por outro lado, a missão institucional do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) é a de fomentar a pesquisa científica, tecnológica e de inovação e contribuir com políticas públicas, para fortalecer o desenvolvimento do Brasil (Apresentação – Conselho Nacional de Desenvolvimento Científico e Tecnológico, [s. d.]). Dessa forma, o CNPq fornece bolsas de iniciação científica, mestrado e doutorado, tanto dentro de projetos científicos quanto como na forma de cotas para programas de pós-graduação, no intuito de fomentar a pesquisa científica. Finalmente, o CNPq, ao diretamente fomentar projetos de pesquisas, indiretamente também contribui para a formação de doutores no Brasil.

No entanto, carecem estudos sobre o impacto do financiamento dessas agências em bolsas de doutorado. Assim, este artigo buscou a análise cientométrica dos currículos daqueles que iniciaram seu doutorado com recorte do ano de 2014, comparando indicadores dos anos 2013 e 2019, buscando identificar o impacto do financiamento de bolsas de doutorado pelo CNPq e pela CAPES. A análise efetuada foi quase-experimental³, por meio de regressão linear multivariada⁴, incluindo variáveis categóricas⁵, e o modelo de diferença em diferenças⁶. Ainda, esse impacto foi comparado ao investimento realizado pelo CNPq, chegando-se à análise custo-efetividade⁷.

3 Na análise experimental, os grupos de tratamento e de controle são escolhidos aleatoriamente. Na quase-experimental, usa-se de artifícios estatísticos para a escolha desses grupos, de modo a reduzir o viés.

4 Regressão com uma variável dependente e duas ou mais independentes. Aqui, a variável dependente é o indicador em 2019, e as múltiplas independentes são o indicador em 2014 e a concessão ou não de bolsas.

5 Variável não numérica, não contínua, que representa uma categoria. Por exemplo, com ou sem bolsas da CAPES.

6 Será detalhado adiante, mas é método que compara os grupos de tratamento e o de controle em dois períodos diferentes para encontrar o impacto da política pública aplicada.

7 Cálculo de quanto custa cada progresso no indicador. Por exemplo, cada artigo publicado custou x reais.

METODOLOGIA

No contexto da ciência aberta, todo o código fonte usado pode ser encontrado no GitHub https://github.com/AlbertoCamposSilva/CV_Lattes_Analyzer. Lá também podem ser encontradas as tabelas necessárias para rodar os códigos aqui indicados, bem como planilhas com os resultados completos das diversas regressões realizadas, retirados todos os identificadores pessoais.

Para a escolha dos dados a serem analisados considerou-se que o objetivo desta pesquisa é encontrar o impacto da concessão de bolsas de doutorado. Como o tempo de doutoramento pode ser de até cinco anos, considerando que a pandemia de 2020 pode causar efeitos imprevisíveis, filtraram-se todos os pesquisadores que iniciaram seu primeiro doutorado no ano de 2014⁸, comparando o currículo de 2013 (antes do seu início) com o de 2019 (após a sua provável conclusão). A base de dados utilizada para a análise cientométrica foi o Currículo Lattes, baixados através do Extrator Lattes⁹.

O Currículo Lattes (Lattes) foi criado em 1990 pelo Governo Federal como forma de padronização dos currículos acadêmicos. O Lattes é obrigatório para os pesquisadores em instituições públicas no Brasil, sendo requisito para a concessão de bolsas e progressões na carreira. Assim, o Lattes se consolidou como fonte de dados para análise bibliométrica (Perlin; Santos, 2015). Para esta pesquisa, foram identificados 7.427.277¹⁰ currículos Lattes. Do total dos currículos, apenas 2.652.105 foram atualizados a partir de 01/01/2020, seleção necessária uma vez que se pretende comparar o ano de 2014 com o de 2019.

Cada um desses currículos foi baixado e tratado de modo a se retirar, deles, indicadores¹¹. Um indicador é aqui referido como um dado, presente no Lattes, com dado temporal definível. Como exemplo, vamos supor que uma determinada pessoa com Lattes atualizado após o ano de 2020 tenha iniciado seu doutorado em educação pela UFRGS em 2014 e concluído em 2018. Desses dados são retirados dois indicadores: início do doutorado (2014) e conclusão do doutorado (2018), sendo ignoradas todas as outras informações referentes a ele, por ausência de ano

8 Código Fonte: 16.2 - Importa ao BD todos os CVs com nível 4 (doutorado), 16.2 Atualiza todos os indicadores de doutorandos 2014.

9 <http://memoria.cnpq.br/web/portal-lattes/extracoes-de-dados>

10 Números Ids Lattes diferentes constantes do arquivo em <http://memoria.cnpq.br/web/portal-lattes/extracoes-de-dados>.

11 Código fonte: Função “atualiza()” em Indicadores.py e 16.2 - Importa ao BD todos os CVs.ipynb.

relacionado. Assim, dos 2.649.498¹² currículos que foram retirados indicadores, encontrou-se um total de 21.250¹³ currículos de pessoas que iniciaram o doutorado em 2014 e o atualizaram em 01/01/2020 ou depois.

Nota-se que os dados apresentados trazem uma limitação na pesquisa, que é a não indicação do impacto do financiamento realizado nas chances de titulação. Isso ocorre porque os dados disponíveis poderiam trazer viés, já que há quem iniciou o doutorado em 2014, mas o fato não está indicado em seu currículo (ou porque esse não foi atualizado, ou porque o doutoramento inconcluso foi excluído dele)¹⁴. Para evitar outro tipo de viés (quem ainda pretende se formar, mas ainda não concluiu o curso), não foram retirados os que não se formaram.

Os currículos foram então separados em quatro distintos grupos: aqueles que foram financiados pela CAPES; os com bolsa do CNPq; os com bolsa de ambas as agências; e os que não mencionam qualquer bolsa recebida¹⁵. O dado sobre o financiamento ou não pela CAPES foi retirado do Lattes, em campo específico em que consta a CAPES como agência financiadora do doutorado, de modo que a exatidão do dado depende de o ex-bolsista ter preenchido corretamente o Lattes¹⁶. O financiamento pelo CNPq foi retirado do portal de dados abertos¹⁷, de modo que se presume ser dado exato e preciso, salvo eventuais homônimos. Deve-se mencionar que, mesmo após tentativas de se contatar a CAPES, não foi possível encontrar dados semelhantes aos fornecidos pelo CNPq.

O estudo então traça os perfis comparativos entre esses grupos, por meio de regressão linear multivariada com variáveis categóricas:

Equação 1 - Medida de um Impacto por meio de Regressão multivariada categórica

$$\text{Indicador}_{2019} = \alpha + \beta \text{Indicador}_{2013} + \gamma \text{CAPES} + \delta \text{CNPq}$$

12 A diferença deste número com o anterior é devido a erros de importação, todos com origem no Extrator Lattes. Infere-se que sejam currículos cancelados ou excluídos.

13 Já excluídos os Ids com erro de importação. SQL: Lista IDs doutorado 2014 para pareamento.sql.

14 Pesquisa informal identificou que muitos doutorandos simplesmente retiram o doutorado inconcluso do currículo, ao invés de o identificar como cancelado, impedindo estudos comparativos entre financiamento e seus efeitos nas chances de conclusão desse.

15 Lista das variáveis categóricas incluídas no estudo: CAPES, CNPq, Ambos e Nenhum.

16 Código Fonte: 16.6 Cria a tabela financiamentos com lista dos IDs que se doutoraram em 2014 e foram ou não financiados.

17 http://dadosabertos.cnpq.br/pt_BR/

Na equação acima, $Indicador_{2013}$ e $Indicador_{2019}$ representam a soma de todas as entradas desse indicador até os anos de referência (2103 e 2019), como, por exemplo, o número total de artigos publicados por um determinado pesquisador até os anos 2013 e 2019, respectivamente. CAPES é uma variável categórica igual a 1, caso o pesquisador tenha recebido bolsas dessa agência, ou caso não tenha, sendo a mesma metodologia aplicada à variável CNPq (sendo os casos de duplo financiamento descartados). Dessa forma, α representa um aumento constante do indicador no tempo; β o aumento temporal desse indicador em relação a seu estado inicial, e aqueles com duplo financiamento foram retirados da pesquisa.

A metodologia é aquela recomendada pelo Banco Mundial em seu livro *Avaliação de Impacto na Prática* (Gertler *et al.*, 2018). Como não é possível método experimental para medir o impacto¹⁸, utiliza-se a segunda melhor opção possível, que é a metodologia quase-experimental. Essa metodologia busca estimar os efeitos do tempo, bem como todos os outros desconhecidos, comparando o estado do Lattes no ano de 2019 com aquele do ano de 2013 de quem não foi financiado, utilizando-se este como grupo de controle.

Assim, a diferença entre os estados dos currículos do grupo de controle com o estado dos currículos nos grupos de tratamento busca eliminar do grupo de tratamento eventual viés decorrente de efeitos externos e dando estimativa do impacto do investimento realizado, o que consiste na metodologia quase-experimental da diferença em diferenças. Nesse caso, essa metodologia é aplicada como modelo da regressão linear já exposta na Equação 1. Assim, $\alpha + \beta Indicador_{2013}$ representam os efeitos no grupo de controle, enquanto $\gamma CAPES + \delta CNPq$ representam uma medida do impacto das bolsas concedidas em cada um dos grupos de tratamento.

A fragilidade da metodologia está na possibilidade de viés no grupo de controle (chamado viés de seleção), ou seja, se a inclinação da reta de progressão no tempo do grupo de controle se diferir da do grupo de tratamento. Um exemplo seria a seleção para o recebimento de bolsas daqueles candidatos com a maior possibilidade de publicar artigos. Nesse caso, a inclinação da reta do grupo de controle seria menor, com menos chances de publicar artigos, que o grupo de tratamento, mesmo sem concessão de bolsas.

Observa-se que, de acordo com Gertler *et al.* (2018), mesmo presente eventual viés, a metodologia usada é a indicada e a capaz de dar os resultados mais próximos da realidade, tendendo a dar resultado subestimado, uma vez que usa de rigor estatístico para desconsiderar indicadores que não apresentaram o nível de confiança exigido.

18 O método experimental significaria sortear quem ganharia bolsa de mestrado e quem não ganharia, algo impraticável por motivos práticos, políticos e legais.

Para buscar diminuir, ou medir, eventuais efeitos do viés acima mencionado, também foi realizada a metodologia do pareamento. De acordo com Gertler *et al.* (2018), uma das formas de se reduzir o viés é buscar alguém com o mesmo perfil (par) e realizar a metodologia acima exposta considerando esse par como o grupo de controle.

Para realizar o pareamento, foi considerada a seleção já indicada (início do doutorado em 2014, atualização do currículo a partir de 2020) e depois o gênero. A seleção pelo gênero se deu pelos diversos estudos existentes indicativos de que mulheres possuem certa desvantagem na ciência em relação aos homens (Areas *et al.*, 2020). Com essas considerações, todos os indicadores do Lattes foram separados em grupos (listados abaixo) e calcularam-se os erros quadrados médios de cada um desses grupos de indicadores, selecionando-se, para cada ex-bolsista do CNPq, o não bolsista, cuja soma dos erros ao quadrado era a menor¹⁹. A amostra foi sem reposição, ou seja, cada doutorando selecionado não poderia ser novamente selecionado, uma vez que a amostragem com reposição verificou forte tendência a que um doutorando fosse selecionado como par de diversos outros. A regressão anteriormente indicada também foi alterada para considerar o valor financiado pelo CNPq. Assim, a equação usada para a regressão foi:

Equação 2 - Medida de Impacto por meio de Regressão multivariada categórica para o CNPq:

$$\text{Indicador}_{2019} = \alpha + \beta \text{Indicador}_{2013} + \theta \text{Financiamento}_{\text{CNPq}}$$

Em que Indicador_{2019} , α , β e Indicador_{2013} mantêm seus significados originais (valor dos indicadores nos respectivos anos, constante e a variável que dá o efeito do indicador em 2013 no indicador em 2019). Mas, $\text{Financiamento}_{\text{CNPq}}$ agora mede o valor financiado²⁰ e θ o impacto desse financiamento. Assim, mede-se mais precisamente o impacto da concessão de bolsas de doutorado pelo CNPq, algo não possível para a CAPES devido à já mencionada falta de dados.

Foi então realizada regressão para cada um dos 140 indicadores, sendo descartadas aquelas com impacto em menos de 1% da população, as não significativas ($P|t| > 0,01$) e as com $R^2_{\text{ajustado}} < 30\%$, método repetido ao longo deste estudo.

19 Código Fonte: 16.7 Pareamento de Doutorados.ipynb.

20 Para melhor precisão e cálculo instantâneo do impacto, a variável sofreu transformada, dividindo-se o montante financiado pela constante do valor total financiado pelo CNPq. A transformada não afeta o resultado final e faz com que a regressão resulte diretamente no impacto total.

R^2_{ajustado} é uma medida que busca indicar o quanto da variância dos dados pode ser explicada pela fórmula encontrada e, portanto, o quanto a fórmula consegue “explicar o mundo real”. R^2_{ajustado} inferior a 0,3 indica, mesmo se houver correlação estatisticamente significativa, que essa é “fraca” e não consegue descrever corretamente a variação dos dados (Grus, 2019). Assim, a representatividade é:

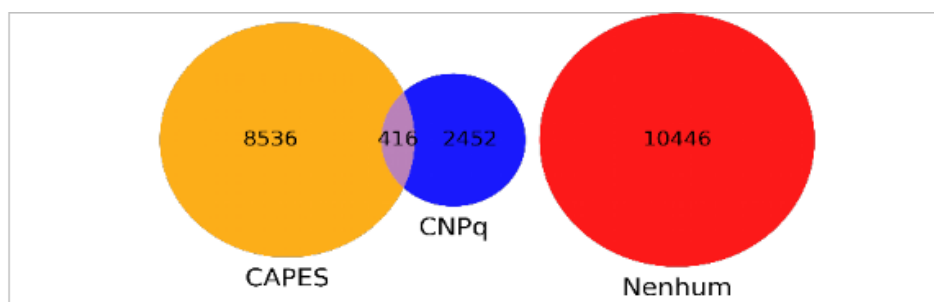
- Fraca: $R^2_{\text{ajustado}} < 0,3$
- Média: $0,3 \leq R^2_{\text{ajustado}} \leq 0,7$
- Forte: $R^2_{\text{ajustado}} > 0,7$

ANÁLISE DOS DADOS

DISTRIBUIÇÃO DOS FINANCIAMENTOS

A primeira análise realizada é em relação ao diagrama de Venn da Figura 1²¹. Nota-se que 52% daqueles que iniciaram seu doutorado receberam bolsas de estudo. O Número está de acordo com a Fundação Nacional de Ciência dos Estados Unidos da América (NSF – National Science Foundation), cuja análise do senso realizado permite inferir que aproximadamente 53% dos doutorandos nas áreas de ciência, engenharia e saúde receberam apoio financeiro na forma de subsídios, bolsas de estudo ou assistência durante seus estudos de pós-graduação. Os dados, tanto nacionais quanto internacionais, demonstram a importância do financiamento a estudantes de bolsas de doutorado. (Doctorate Recipients from U.S. Universities: 2021 | NSF - National Science Foundation, [s. d.])

Figura 1 - Distribuição da concessão de bolsas de doutorado daqueles que iniciaram seu doutorado em 2014



Fonte: autoria própria, com base no Lattes e no portal de dados abertos do CNPq (2023).

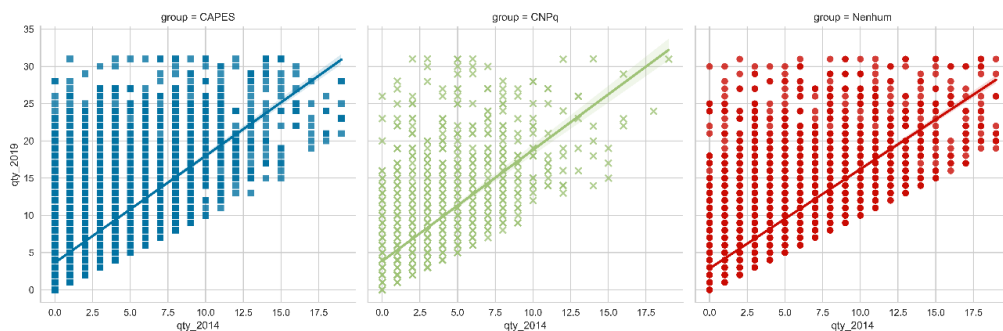
21 Código Fonte: 15.5 Diagrama de Venn.

A segunda conclusão a que se chega do mesmo diagrama é que o CNPq financiou apenas 13% dos doutorandos, enquanto a CAPES concedeu bolsas a 41%. Dessa forma, nota-se que a presença da CAPES nesse assunto é maior que a do CNPq.

DISTRIBUIÇÃO DOS DADOS

Um dos pressupostos mais usuais em ciência de dados e também um dos que menos são verdadeiros é o de que os dados (ou a distribuição dos erros em caso de regressão linear) estejam normalmente distribuídos. As implicações da não normalidade são diversas e podem depender de cada caso (Grus, 2019), tornando o estudo necessário. Quanto à normalidade dos dados, usou-se o teste de Shapiro Wilk, em que se encontrou a não normalidade (p -valor menor que 0,01). A distribuição do indicador Artigos Publicados consta da Figura 2.

Figura 2 - Distribuição do indicador Artigos Publicados de acordo com a concessão de bolsas



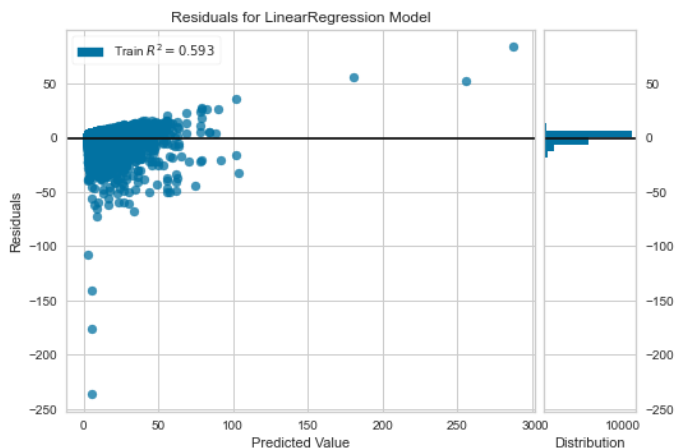
Fonte: Autoria própria (2023).

Para verificação matemática da distribuição dos erros nas regressões realizadas, verificou-se, nas regressões realizadas, que o “Prob(Omnibus)”, que é o teste que busca verificar a hipótese nula de que a distribuição dos erros é normal, foi normalmente inferior a 0.000. Ou seja, temos mais de 99,9% de chance de que os erros não estão normalmente distribuídos.

Tentou-se então várias transformadas para a busca de uma distribuição normal, sendo o melhor resultado obtido com as seguintes transformadas: remoção dos *outliers* (pontos acima de quatro vezes o desvio-padrão), aplicação da fórmula $\sqrt{\log(\text{indicador})}$ nas somas tanto de 2019 quanto na de 2013, e remoção de todos os pontos NaN (*not a number*). Mesmo com essas transformadas, encontrou-se que

a distribuição dos erros não era próxima à curva normal (Prob(Omnibus): 0.000) e verificou-se que R^2 diminuiu significativamente. Dessa forma, optou-se por não realizar nenhuma transformada ou exclusão de *outliers*.

Figura 3 - Distribuição dos Erros da Regressão Linear do Indicador Artigos Publicados



Fonte: Autoria própria (2023)

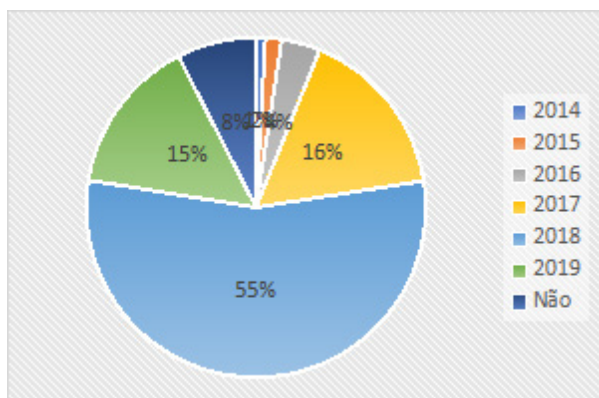
A principal consequência da não normalidade dos erros está na capacidade preditiva da equação, mais sensibilidade a *outliers*, redução da precisão do teste de significância e da precisão em grupos de tamanhos distintos. Mesmo assim, como há muitos dados disponíveis, o teorema do limite central indica a normalidade assintótica e a correta aplicação da ferramenta estatística (Bernier; Feng; Azakawa, 2011 *apud* Alberto Pino; Paulo, [s. d.]), principalmente porque não se buscam modelos preditivos. Além disso, observa-se que o uso da constante α é recomendada devido à não normalidade dos erros (Pino, 2014).

A regressão linear é, portanto, ferramenta indicada para que se chegue ao mínimo dos erros quadráticos e a melhor ferramenta para dar o impacto médio da concessão de bolsas.

ANO DE CONCLUSÃO DO DOUTORADO

Analizando a relação entre o financiamento e o tempo de conclusão do doutorado, temos a seguinte distribuição:

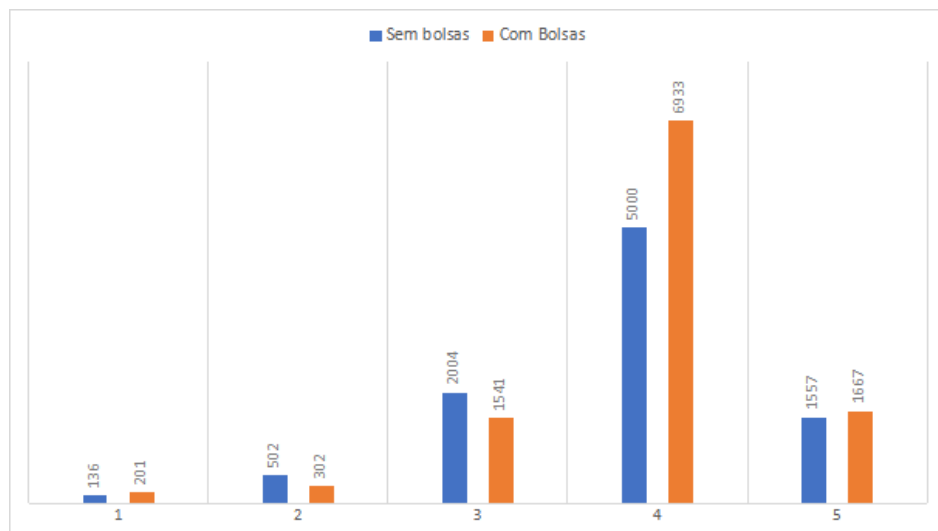
Figura 4 - Ano de conclusão do doutorado iniciado em 2014 (%).



Fonte: autoria própria (2023).

A Figura 4 apresenta que 55% das pessoas que iniciaram seu doutorado em 2014 o concluíram em 4 anos, sendo que apenas 8% da população não concluiu o doutorado até o ano de 2019.

Figura 5 - Comparação do tempo de doutoramento - com e sem bolsas de estudo.



Fonte: Autoria própria, com base no Lattes e no portal de dados abertos do CNPq (2023).

Uma questão que pode ser levantada é se a concessão de bolsas afeta esse tempo de conclusão em alguma forma. Para o levantamento, os doutorandos foram separados em dois grupos distintos: com bolsas e sem bolsas. Assim, foi realizada regressão linear com o seguinte modelo:

Equação 3 - Financiamento influencia no tempo de doutoramento?

$$Y_i = \alpha + \beta X_i$$

Na equação acima, Y_i é a subtração do ano de conclusão do doutorado com o ano de início deste (aqui, 2014); α é a constante e representa o tempo médio para que se conclua o doutorado; X é variável categórica de valor 1, caso o doutorando tenha recebido bolsa de estudos, ou zero, caso não tenha; de modo que β mede a influência, em anos, do financiamento no tempo de doutoramento. Encontrou-se que:

- $\alpha = 3.8$ (tempo médio de doutoramento), com $P > |t| = 0,000$ (significativo);
- $\beta = 0.086$ (ser financiado aumenta em um mês o tempo de doutoramento), com $P > |t| = 0,000$ (significativo);
- $R^2_{\text{ajust.}} = 0.002$ (ser financiado influencia apenas em 0,2% o tempo de doutoramento); e
- Prob (F-statistic): $3.83e-12$ (a regressão é estatisticamente relevante).

Conclui-se que, embora haja correlação entre a concessão de bolsas e o aumento em um mês do tempo para a titulação, tanto o valor de β quanto o de $R^2_{\text{ajust.}}$ são o suficientemente pequenos para podermos concluir que não há impacto da concessão de bolsas no tempo necessário para concluir o doutorado. Vale mencionar que esse mesmo raciocínio será adotado em outras análises abaixo.

Também pode-se concluir que o tempo médio de doutoramento é de cerca de quatro anos e que a janela de cinco anos inclui ao menos 92% dos que se doutoraram, de forma a ser uma janela temporal estatisticamente conclusiva.

IMPACTO DO FINANCIAMENTO

Foi montada tabela com os IDs de todos os currículos que tiveram o início de seu primeiro doutorado em 2014, resultando em 21.250 linhas²². Tabela com os indicadores foi gerada²³ e filtrada por indicador²⁴, para que então a regressão linear fosse gerada, sendo usada a seguinte fórmula:

Equação 4 - Medida de um Impacto por meio de Regressão multivariada categórica

$$\text{Indicador}_{2019} = \alpha + \beta \text{Indicador}_{2013} + \gamma \text{CAPES} + \delta \text{CNPq}$$

Assim, a regressão linear foi realizada em cada um dos 140 indicadores²⁵. Dos resultados, retiraram-se aqueles cujos Prob (F-statistic) > 0.01 (mais de 1% de chance de que a regressão não é estatisticamente relevante), os com impacto pequeno (menor que 212), e os indicadores cujos $R^2_{\text{ajustado}} < 0,3$. Assim, restaram os seguintes indicadores:

22 Tabela: ids_doutorandos_inicio_em_2014. SQL: Cria tabela com listas de IDs que iniciaram o doutorado em 2014.sql.

23 Código Fonte: 16.2 - Importa ao BD todos os CVs.

24 SQL: Cria tabela de indicadores para doutorado.sql.

25 Código Fonte: 16.7 Fazendo Regressão CAPES, CNPQ, Nenhum.

Tabela 1 - Regressão Linear com Medida de um Impacto de Concessão de Bolsas de Doutorado

Indicador	Nº	R ² _{ajustado}	α	β	γ	δ	$\frac{\gamma+\delta}{2}$
Artigo Publicado	19173	46%	3,6	1,58	1,39	1,81	1,60
Início de Vínculo Empregatício ou Funcional	20671	67%	1,66	1,29	1,46	1,66	1,56
Início de Atividade de Ensino	9463	72%	2,08	1,24	1,41	1,33	1,37
Fim de Atividade de Ensino	8131	73%	2,21	1,26	1,11	1,09	1,10
Produção de Carta, Mapa ou Similar	252	71%	0	1,22	1,96	0	0,98
Trabalho em Eventos	18283	74%	3,95	1,35	0,95	0,87	0,91
Detalhamento do Trabalho	18248	74%	3,9	1,35	0,91	0,79	0,85
Fim de Vínculo Empregatício ou Funcional	20025	73%	1,5	1,22	0,87	0,74	0,81
Ministério de Curso de Curta Duração	11061	75%	1,03	1,31	0,43	0,52	0,48
Produção de Trabalho Técnico	7032	55%	3,33	1,2	0,77	0	0,39
Conclusão de Curso de Curta Duração	16545	77%	1,63	1,17	0,26	0,36	0,31
Início de Curso de Curta Duração	16568	77%	1,64	1,17	0,26	0,36	0,31
Fim de Atividade de Pesquisa e Desenvolvimento	3505	66%	0,75	0,93	0,17	0,28	0,23
Apresentação de Trabalho Técnico	18894	72%	3,82	1,38	0,42	0	0,21
Início de Estágio	4132	83%	0,23	1	0,21	0,17	0,19
Fim de Estágio	3968	83%	0,23	1	0,21	0,16	0,19
Início de Atividade de Pesquisa e Desenvolvimento	7093	61%	0,83	0,96	0,13	0,15	0,14
Participação em Congresso	18769	79%	1,86	1,22	0	0,24	0,12
Prêmio ou Título Recebido	11599	67%	1,24	1,17	0,13	0	0,07
Participação em Simpósio	14864	74%	0,98	1,13	0,09	0	0,05
Início de Participação em Projeto	16636	70%	1,6	1,23	-0,12	0	-0,06
Início de Projeto de Pesquisa	16537	61%	2,01	1,3	-0,18	0	-0,09
Publicação de Capítulo de Livro	11476	43%	1,95	1,42	0	-0,27	-0,14
Artigo Aceito para Publicação	3890	38%	0,6	1,03	-0,12	-0,15	-0,14
Fim de Projeto de Pesquisa	14815	57%	1,86	1,24	-0,17	-0,24	-0,21
Outras Orientações Concluídas	12780	71%	7,53	1,28	-1,33	-1,92	-1,63
Participação em Banca de Graduação	13381	72%	7,19	1,24	-1,19	-2,16	-1,68

Fonte: Autoria própria, com base no Lattes e no portal de dados abertos do CNPq (2023).

Com base na tabela acima, podemos agrupar as atividades com “início” e “fim”, subtraindo o impacto encontrado em “início” daquele que indica o “fim” para se chegar ao quantitativo de uma “Nova” entrada, bem como calcular o impacto. Assim, a seguinte tabela pode ser construída²⁶:

Tabela 2 - Impacto de Concessão de Bolsas de Doutorado pelo CNPq e pela CAPES.

Indicador	Nº CAPES	Impacto CAPES	Nº CNPq	Impacto CNPq	Impacto Total
Artigo Publicado	7833	10856	2297	4163	15.019
Trabalho em Eventos	7493	7084	2122	1844	8.927
Detalhamento do Trabalho	7476	6790	2118	1683	8.473
Novo Vínculo Empregatício ou Funcional	142	4951	57	2247	7.198
Apresentação de Trabalho Técnico	7653	3251	2183	0	3.251
Conclusão de Curso de Curta Duração	6770	1764	1851	671	2.435
Ministério de Curso de Curta Duração	4243	1807	1103	577	2.384
Produção de Trabalho Técnico	2675	2060	669	0	2.060
Nova Atividade de Ensino	404	1454	117	353	1.807
Prêmio ou Título Recebido	4521	586	1416	0	586
Participação em Simpósio	6098	535	1754	0	535
Participação em Congresso	7609	0	2210	531	531
Novo Projeto de Pesquisa	685	-176	186	428	253
Produção de Carta, Mapa ou Similar	117	229	26	0	229
Nova Atividade de Pesquisa e Desenvolvimento	1341	125	330	-6	119
Novo Estágio	60	1	11	7	8
Artigo Aceito para Publicação	1561	-192	421	-65	-257
Publicação de Capítulo de Livro	4432	0	1156	-312	-312
Outras Orientações Concluídas	4721	-6265	1233	-2372	-8.637
Participação em Banca de Graduação	5307	-6330	1325	-2856	-9.185

Fonte: Autoria própria, com base no Lattes e no portal de dados abertos do CNPq (2023).

Dos dados encontrados acima, analisaremos com mais cuidado apenas dois exemplos, de modo que a análise efetuada possa ser facilmente estendida aos outros casos.

ARTIGOS PUBLICADOS

O R^2_{ajustado} do indicador relacionado a Artigos Publicados é de 46%, ou seja, a fórmula aplicada consegue explicar cerca de 46% da variação do quantitativo de artigos publicados até 2019. Como a estatística é significativa (o maior valor de $P > |t|$ foi de 4-10-30, em δ), temos a ida (forte correlação estatística) e a volta (o modelo adotado descreve mediantemente os dados observados), de forma que podemos inferir que houve impacto. Podemos assim interpretar os números:

- α : para se doutorar, o pesquisador publica em média 3,6 artigos.
- β : já ter publicado antes de iniciar seu doutorado se correlaciona com mais 1,6 artigos publicados por artigo inicial e até o ano de 2019.
- γ : ter recebido bolsas da CAPES se correlaciona com mais 1,4 artigos publicados até o ano de 2019.
- δ : ter recebido bolsas do CNPq se correlaciona com mais 1,8 artigos publicados até o ano de 2019.

A amostra contempla apenas 19.173 dos 21.250 doutorandos analisados, o que indica que 1.714 doutorandos não publicaram artigos até o ano de 2019, não os incluíram em seu Lattes ou foram excluídos por terem recebido bolsas de ambas as agências (416 casos).

VÍNCULOS

Há 15.948 doutorandos que iniciaram algum tipo de vínculo (com instituições) até o ano de 2019, e 8.750 que encerraram qualquer tipo de vínculo. Examinados em separados, esses números são pouco significativos, assim como qualquer indicador que tenha “Início” e “Fim” em seus nomes. Mas, examinados em conjunto, podemos verificar que houve 7.198 novos vínculos firmados devido ao doutorado.

Vínculo é uma entrada no Lattes, de forma que qualquer alteração desse campo no Lattes é contabilizada. Ou seja, os números não indicam 7.198 novos empregos, mas alterações nos vínculos, que podem ser, por exemplo, uma mudança de lotação, uma promoção, um novo emprego ou mesmo o desemprego.

Como exemplo, há casos de vínculos “bolsista da CAPES”, o que explica parte dos 4.951 novos vínculos devido aos financiamentos pela CAPES. Por outro lado, verificou-se a falta de atualização de alguns currículos quando do término da bolsa da CAPES, de modo que o bolsista registra o início dessa, mas não seu término.

Finalmente, analisando apenas o grupo de controle:

Equação 5 - Impacto Negativo em Novos Vínculos

$$Indicador_{2019} = \alpha + \beta Indicador_{2013}$$

$$Impacto = [(1,66 * 20.671) - (1,50 * 20.025)] = -4.235$$

O impacto negativo de 19% dá indicativo de desistência de vínculos durante o processo de titulação para quem não recebeu bolsas, algo que pode ser melhor analisado.

PAREAMENTO

Para análise do Impacto da concessão de bolsas pelo CNPq, pode-se adotar metodologia em que se escolhe, para cada pessoa tratada, uma pessoa semelhante para compor o grupo de tratamento, em busca de eliminar possíveis vieses de seleção (i.e., quando a política de concessão escolhe o grupo tratado por variáveis capazes de influenciar no resultado) (Gertler *et al.*, 2018). Dessa forma, foi realizado o processo de pareamento²⁷, conforme metodologia indicada pelo livro do banco mundial *Avaliação de Impacto na Prática* (Gertler *et al.*, 2018). Assim, para cada doutorando que recebeu financiamento do CNPq, foi escolhido um par da seguinte forma:

- Iniciou seu doutorado em 2014,
- Atualizou seu Lattes a partir de 1/1/2020,
- Do mesmo sexo,

27

16.7 Pareamento de Doutorados.

• Com a minimização de erros quadráticos nos seguintes grupos de indicadores:

- Atividades de Pesquisa e Desenvolvimento,
- Livros Publicados,
- Atividades de Ensino,
- Publicações de Artigos,
- Vínculos Empregatícios ou Funcionais,
- Mestrados,
- Patentes e
- Participação em Congressos.

Para a minimização dos erros quadráticos, para cada doutorando financiado pelo CNPq, calculou-se a diferença da soma dos indicadores em cada grupo, ao quadrado. Assim, por exemplo, se um doutorando teve dois livros publicados, e outro nenhum, o erro de “Livros Publicados” é de $(2-0)^2 = 4$. Somando-se esse erro para todos os outros grupos, temos a soma dos erros quadráticos.

Uma vez realizado o pareamento, pode-se aplicar o método da diferença em diferenças calculando a diferença dos indicadores no grupo de tratamento com os do grupo de controle. Ou seja:

Equação 6 - Diferença em Diferenças

$$\text{Diferença em Diferenças} = [\text{Estado}_{\text{atual}} - \text{Estado}_{\text{inicial}}]_{\text{Grupo de Tratamento}} - [\text{Estado}_{\text{atual}} - \text{Estado}_{\text{inicial}}]_{\text{Grupo de Controle}}$$

A equação acima, levada para nosso caso em questão, resulta em:

Equação 7 - Cálculo do Impacto pelo método da Diferença em Diferenças

$$\text{Impacto} = [\text{Indicador}_{2019} - \text{Indicador}_{2013}]_{\text{Com Bolsas do CNPq}} - [\text{Indicador}_{2019} - \text{Indicador}_{2013}]_{\text{Sem Bolsas do CNPq}}$$

Aplicando a fórmula acima ao Banco de Dados²⁸, temos o seguinte impacto²⁹, sendo aqui listadas apenas as 10 primeiras e as 3 últimas linhas, ordenadas pelo Impacto:

28 Código Fonte: Impacto do Financiamento pelo cálculo simples de Diferença em Diferenças.sql.

29 Tabela completa: Impacto por meio da Diferença em Diferenças simples.xlsx.

Tabela 3 - Diferença em Diferenças por Pareamento dos Doutorandos 2014 por soma simples.

Indicador	Sem Financiamento			Financiado pelo CNPq			Impacto
	2019	2013	Diferença	2019	2013	Diferença	
Trabalho em Eventos	22451	13012	9439	40688	22857	17831	8392
Artigo Publicado	10446	2991	7455	18442	4165	14277	6822
Início de Vínculo Empregatício ou Funcional	19758	13046	6712	23755	12612	11143	4431
Apresentação de Trabalho Técnico	20253	10959	9294	27886	14897	12989	3695
Início de Projeto de Pesquisa	5786	3198	2588	11184	5991	5193	2605
Participação em Congresso	10990	7136	3854	14957	8677	6280	2426
Fim de Vínculo Empregatício ou Funcional	14350	9544	4806	17351	10190	7161	2355
Início de Participação em Projeto	4923	2877	2046	9583	5532	4051	2005
Fim de Projeto de Pesquisa	3975	2143	1832	7740	4048	3692	1860
Produção de Trabalho Técnico	3068	1869	1199	4506	1635	2871	1672
(...)							
Música	1728	1240	488	430	248	182	-306
Outras Orientações Concluídas	15436	6913	8523	10760	3178	7582	-941
Participação em Banca de Graduação	15346	6480	8866	10048	3041	7007	-1859

Fonte: Autoria própria, com base no Lattes e no portal de dados abertos do CNPq (2023).

Exemplificando, a soma de todos os artigos publicados para quem não recebeu bolsas até o ano de 2013 foi de 2.991, enquanto a soma em 2019 foi de 10.446. Portanto, a diferença entre os anos de 2014 e 2019 indica que esse grupo publicou $10.446 - 2.991 = 7.455$ artigos. Já o grupo que recebeu bolsas do CNPq publicou, calculando as diferenças de publicações nesse mesmo período, $18.442 - 4.165 = 14.277$ artigos. Assim, a diferença dessas diferenças é $14.277 - 7.455 = 6.822$, o que indica que a concessão de bolsas a 2.296 pesquisadores teve um impacto de uma média de 3 artigos a mais por pesquisador. Essa metodologia é conhecida como “Diferença em Diferenças”.

Esse método de cálculo é simples e útil, mas não tem o poder estatístico de uma regressão linear. Assim, como o CNPq possui informações sobre o montante financiado em bolsas de doutorado para cada doutorando, adotou-se a metodologia de econometria de regressão linear, usando o modelo de diferença em diferenças e que buscou calcular o efeito desse financiamento, de modo que a Equação 1 foi alterada para:

Equação 8 - Medida do Impacto do CNPq por meio de Regressão multivariada

$$Indicador_{2019} = \alpha + \beta Indicador_{2013} + \theta Financiamento_{CNPq}$$

Na equação acima, $Indicador_{2013}$, $Indicador_{2019}$, α e β mantêm o significado original (soma dos indicadores até os respectivos anos, constante e influência de um indicador no outro com o tempo). $Financiamento_{CNPq}$ dá a soma total investida pelo CNPq ao bolsista em bolsas de doutorado, tanto no país quanto no exterior, entre os anos de 2014 e 2019. Assim, θ dá a medida de impacto desse financiamento, de modo que $\theta Financiamento_{CNPq}$ dá o efeito desse em $Indicador_{2019}$, sendo, portanto, medida de efetividade.

Realizando as diversas regressões acima³⁰, filtrando-se os indicadores não significativos (Prob (F-statistic) < 0,01; $P > |t| < 0,01$, amostra > 200, $R^2_{ajust} > 0.3$), podemos chegar à seguinte tabela:

30 Código Fonte: 17.1 - Regressões Lineares - CNPq e Pagtos, Sem preencher nem best_r2.ipynb.

Tabela 4 - Impacto de Bolsas do CNPq aos Doutorandos 2014³¹.

Indicador	Nº	R ² ajustado	α	β	Impacto	Erro Padrão (+/-)
Artigo Publicado	4148	45%	3,43	1,93	1924	339
Início de Vínculo Empregatício ou Funcional	4695	61%	1,98	1,29	1922	241
Fim de Vínculo Empregatício ou Funcional	4494	70%	1,79	1,19	643	165
Participação em Congresso	3996	69%	1,75	1,20	586	194
Participação em Encontro	3334	77%	0,71	1,17	367	140
Início de Residência Médica	142	99%		1,00	239	66
Orientação em Andamento de Graduação	436	34%	1,13	1,15	-1034	386
Participação em Banca de Graduação	2561	54%	6,39	1,18	-3274	950
Outras Orientações Concluídas	2414	50%	7,22	1,10	-3438	1018

Fonte: Autoria própria com base no Portal de Dados Abertos do CNPq (2022).

Observa-se que a Tabela 6 apresenta impacto medido muito menor que o da Tabela 3. Pode-se considerar que a Tabela 3 dá medida de impacto sem descarte estatístico, enquanto a Tabela 6 dá apenas o impacto com alta probabilidade estatística. Já a Tabela 2 dá o Impacto sem levar em conta o valor financiado. Dessa forma, o impacto real deve estar entre essas tabelas, considerando-se o da Tabela 3 sobrestimado e o da Tabela 6 subestimado, mas que, provavelmente, o modelo mais próximo da realidade seja o da Tabela 3.

ANÁLISE CUSTO-EFETIVIDADE

A Análise custo-efetividade, que é muito aplicada em economia e bastante utilizada para permitir a comparação entre distintas políticas públicas, busca comparar aquilo que se obteve com o custo de se obter esse resultado. Considerando que o custo é o total investido pelo CNPq em bolsas de doutorado, é constante no valor de R\$ 289.924.589,50, de modo que podemos dividir esse valor pelo impacto medido para ter a medida de custo-efetividade. Análise semelhante não pôde ser realizada com a CAPES pois essa não forneceu informações financeiras.

31 Resultado completo em Regressões variáveis_independentes ['qty_2014'^J 'porcent_pagos'] confianças [0.05] 2014 Pareamento False best_r2False CNPq True CAPES False Ambos False Nenhum True preenche_todos_idsFalse.xlsx.

Utilizando a mesma metodologia anterior ($\text{Prob (F-statistic)} < 0,01$; $P > |t| < 0,01$, amostra > 200 , $R^2_{\text{ajust}} > 0,3$) e unindo indicadores com “Início” e “Fim” para que seja construído o indicador “Novo”:

Tabela 5 - Custo-efetividade de bolsas de doutorado pelo CNPq a quem iniciou seu doutorado em 2014.

Indicador	Impacto	Custo-Efetividade
Participação em Banca de Graduação	-2768	-R\$ 96.240,32
Outras Orientações Concluídas	-2945	-R\$ 90.475,31
Artigo Publicado	2053	R\$ 129.746,04
Tradução	1978	R\$ 134.663,90
Novo Vínculo Empregatício ou Funcional	1294	R\$ 205.887,89
Participação em Exposição	664	R\$ 401.391,35
Participação em Banca de Mestrado	-643	-R\$ 414.532,95
Conclusão de Aperfeiçoamento	441	R\$ 604.521,22
Nova Atividade de Ensino	363	R\$ 734.066,58
Participação em Oficina	265	R\$ 1.005.679,58
Início de Residência Médica	200	R\$ 1.331.548,51

Fonte: Autoria própria (2023).

Finalmente, podemos comparar os resultados anteriores (em que foi verificando se houve ou não bolsas pela CAPES ou pelo CNPq – Tabela 2), aos resultados da Tabela 5, chegando-se à seguinte tabela:

Indicador	Custo-Efetividade por Concessão ou não	Custo-Efetividade por Pareamento	Média
Participação em Banca de Graduação	-R\$ 101.514,21	-R\$ 88.553,63	-R\$ 95.033,92
Outras Orientações Concluídas	-R\$ 122.227,90	-R\$ 84.329,43	-R\$ 103.278,67
Artigo Publicado	R\$ 69.643,19	R\$ 150.688,46	R\$ 110.165,82
Novo Vínculo Empregatício ou Funcional	R\$ 129.027,41	R\$ 226.680,68	R\$ 177.854,04
Orientação em Andamento de Aperfeiçoamento ou Especialização	#N/D	-R\$ 133.667,40	-R\$ 133.667,40
Trabalho em Eventos	R\$ 157.225,92	#N/D	R\$ 157.225,92
Detalhamento do Trabalho	R\$ 172.266,54	#N/D	R\$ 172.266,54
Orientação em Andamento de Graduação	#N/D	-R\$ 280.391,29	-R\$ 280.391,29
Conclusão de Curso de Curta Duração	R\$ 432.078,37	#N/D	R\$ 432.078,37
Início de Curso de Curta Duração	R\$ 435.322,21	#N/D	R\$ 435.322,21
Ministério de Curso de Curta Duração	R\$ 502.468,96	#N/D	R\$ 502.468,96
Participação em Congresso	R\$ 545.997,34	R\$ 494.751,86	R\$ 520.374,60
Novo Projeto de Pesquisa	R\$ 677.393,90	#N/D	R\$ 677.393,90
Participação em Encontro	#N/D	R\$ 789.985,26	R\$ 789.985,26
Nova Atividade de Ensino	R\$ 821.316,12	#N/D	R\$ 821.316,12
Publicação de Capítulo de Livro	-R\$ 929.245,48	#N/D	-R\$ 929.245,48
Início de Residência Médica	#N/D	R\$ 1.213.073,60	R\$ 1.213.073,60

Fonte: Autoria própria (2023).

Conforme já explicado, a regressão por pareamento busca eliminar possível viés de seleção e, por isso, é mais precisa. Dessa forma, como medida de custo-efetividade, deve-se considerar, preferencialmente, essa forma de regressão em relação à da Tabela 2. De qualquer forma, observa-se que as medidas, quando disponíveis em ambas as formas, não se afastam muito uma da outra, o que comprova a robustez do método aplicado.

CONCLUSÃO

O estudo confirmou a importância da concessão de bolsas para a formação no nível de doutorado. Ainda, a análise cientométrica realizada deu medida estatística do impacto da concessão de bolsas, tanto pela CAPES quanto pelo CNPq, de três distintas formas.

Na Tabela 6, os dois primeiros impactos são negativos. Mas, ambos são impactos que se traduzem em atividades de pouca especialização: Participação em Banca de Graduação e Outras Orientações Concluídas. Em seguida, vemos Artigos Publicados, com a custo-efetividade de R\$ 110.165,82 por artigo. O custo por artigo não é alto se pensarmos que um artigo significa um avanço na fronteira do conhecimento e a introdução de mais um pesquisador brasileiro. Ainda, a bolsa de doutorado do CNPq possui o valor de R\$ 148.800,00 (R\$ 3.100,00 mensais à época), o que se traduz num impacto não só de um artigo por doutor, mas de um novo vínculo e um trabalho publicado em evento por doutorando.

O estudo é capaz de indicar impactos negativos, como a publicação de capítulos de livros, de modo a indicar pontos que devem ser incentivados aos doutorandos. Por exemplo, pode-se incentivar que doutorandos orientem graduandos ou que também publiquem livros, ao invés de só artigos, de modo a reduzir esse impacto negativo e possibilitar uma formação mais ampla por parte do doutorando.

O estudo permite a comparação entre diversas políticas públicas dando indicação de qual política possui uma efetividade melhor e, portanto, servindo como orientação. Alerta-se, no entanto, que sem uma análise mais detalhada a metodologia não consegue indicar qual é a “melhor” política pública. Aliás, essa é justamente a grande vantagem, mas também a grande limitação do estudo: por ser quantitativo, é escalável a diversas políticas públicas, facilmente aplicado uma vez desenvolvido e comparável no tempo, mas também incapaz de dar um *insight* mais aprofundado e indicar os mecanismos por trás desse impacto (ou da falta dele), bem como as consequências da política pública de forma mais ampla.

Por isso, sugere-se continuidade da pesquisa a outras políticas públicas e também na forma de séries temporais, bem como estudos de caso que poderão confirmar o impacto e indicar as mudanças sociais decorrentes desse. Além disso, seria interessante confirmar quais das medidas de impacto é a mais precisa.

Um estudo de caso em potencial que já se sugere é a correlação entre doutorado e novos vínculos. O impacto negativo no grupo de controle sugere a desistência do vínculo atual. Por outro lado, o forte impacto positivo em geral sugere que as bolsas de estudo são um excelente remédio ao problema.

Em consideração final, os investimentos praticados, tanto pela CAPES quanto pelo CNPq, no geral, têm impacto positivo e pode-se dizer que proporcionaram, aproximadamente, a publicação de 15.019 artigos, 8.927 trabalhos em eventos, 7.198 novos vínculos, 2.060 trabalhos técnicos, 1.807 novas atividades de ensino, entre diversos outros, o que colabora com as metas do PNE e com o desenvolvimento socioeconômico do país.

REFERÊNCIAS

APRESENTAÇÃO — CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO. [S. l.], [s. d.]. Disponível em: <https://www.gov.br/cnpq/pt-br/acesso-a-informacao/institucional/institucional>. Acesso em: 5 abr. 2023.

AREAS, R. *et al.* **Gender and the scissors graph of Brazilian science: from equality to invisibility.** [s. l.], 2020. Disponível em: <https://osf.io/m6eb4/download>. Acesso em: 28 abr. 2023.

DOCTORATE RECIPIENTS FROM U.S. UNIVERSITIES: 2021 | NSF - NATIONAL SCIENCE FOUNDATION. [S. l.], [s. d.]. Disponível em: <https://ncses.nsf.gov/pubs/nsf23300/data-tables>. Acesso em: 7 abr. 2023.

GERTLER, P. J. *et al.* **Avaliação de impacto na prática.** Segunda Eded. [S. l.]: World Bank Publications, 2018. E-book. Disponível em: <https://openknowledge.worldbank.org/bitstream/handle/10986/25030/9781464808890.pdf>. Acesso em: 21 jun. 2019.

GRUS, J. **Data science from scratch: first principles with python.** [s. l.], 2019. Disponível em: https://books.google.com.br/books?hl=pt-BR&lr=&id=YBKSDwAAQBAJ&oi=fnd&pg=PP1&dq=python+data+science+book&ots=bKcGyQvrvG&sig=8mW_IcLnO_WbscWKcCTga5mqAY. Acesso em: 28 abr. 2023.

HISTÓRIA E MISSÃO — CAPES. [S. l.], [s. d.]. Disponível em: <https://www.gov.br/capes/pt-br/acesso-a-informacao/institucional/historia-e-missao>. Acesso em: 5 abr. 2023.

PERLIN, M. S.; SANTOS, A. A. P. **Os pesquisadores, as publicações e os periódicos da área de Finanças no Brasil: Uma análise com base em currículos da plataforma Lattes.** Revista brasileira de finanças, [s. l.], v. 13, n. 2, p. 162–199, 2015. Disponível em: <http://hdl.handle.net/10183/130096>. Acesso em: 12 mar. 2023.

PINO, F. A. **A Questão da Não Normalidade: uma Revisão.** Revista de economia agrícola, [s. l.], v. 61, n. 2, p. 17–33, 2014. Disponível em: <http://www.iea.sp.gov.br/ftpica/publicar/rea2014-2/rea2-22014.pdf>. Acesso em: 25 abr. 2023.

PNE - PLANO NACIONAL DE EDUCAÇÃO - PLANO NACIONAL DE EDUCAÇÃO - PNE. [S. l.], [s. d.]. Disponível em: <https://pne.mec.gov.br/>. Acesso em: 5 abr. 2023.

PRESIDÊNCIA DA REPÚBLICA; MINISTÉRIO DA EDUCAÇÃO; SECRETARIA DE ARTICULAÇÃO COM OS SISTEMAS DE ENSINO. **Planejando a Próxima Década.** [S. l.: s. n.], 2014. E-book. Disponível em: https://pne.mec.gov.br/images/pdf/pne_conhecendo_20_metas.pdf. Acesso em: 5 abr. 2023.

SILVA, V. C. da; PORCIÚNCULA, L. de O. **Impactos da pós-graduação stricto sensu na formação de professores de português da educação básica do Distrito Federal.** Formação Docente – Revista Brasileira de Pesquisa sobre Formação de Professores, [s. l.], v. 9, n. 17, p. 141–162, 2017. Disponível em: <https://www.revformacaodocente.com.br/index.php/rbpf/article/view/156>. Acesso em: 17 abr. 2023.

SOBRE OS AUTORES

Alberto de Campos e Silva

Formado em física pela UnB, atuação como professor do ensino fundamental, gerente e programador do Cartório Distribuidor Cível da Comarca de Goiânia. Atualmente Analista de C&T e assessor da Diretoria Científica do CNPq. Especialização em Gestão Governamental, Gestão Pública, Ciência de Dados e Big Data Analytics, Mestrado em Gestão Econômica do Meio Ambiente e Doutorado em Educação em Ciências. Apesar dos títulos do mestrado e do doutorado, esses são, na verdade, o uso da cientometria (indicadores de CT&I) para a análise do impacto no CNPq e, portanto, se correlacionam com gestão pública, ciência de dados e big data. Cursando MBA em Gestão de Pessoas e em Data Science e Analytics, pela USP. Além disso, formação complementar em Projetos, Processos, Propriedade Intelectual, Impacto de Políticas Públicas, Plano Estratégico, Avaliação de Banco de Dados, Avaliação de Ações e Avaliação de CT&I. Experiência com programação (especialmente Python) e banco de dados relacional. Assessor da Diretoria Científica do CNPq, com participação em diversos Grupos de Trabalho para as mais diversas funções, trabalhando ativamente na gestão da área finalística do órgão. E-mail: alberto@albertocampos.com.br

Alexandre Guilherme Motta Sarmiento

Orientador de Doutorado e Mestrado. Sociólogo com graduação pela Universidade de Brasília; Especialização e Mestrado em Sociologia pela mesma Universidade; é Doutor em Educação em Ciência pela Universidade Federal de Santa Maria. Fez pós-doutorado no PPG em Ciências da Computação da Universidade Estadual de Maringá (Manna Pesquisa, Ensino e Inovação). É Analista em Ciência e Tecnologia do Conselho Nacional de Desenvolvimento Científico e Tecnológico; Professor e Pesquisador no PPG em Educação e Ciências da UFRGS, Pesquisador e Coordenador do Grupo de Pesquisa em Ciência, Tecnologia e Sociedade: Ensino, Pesquisa e Desenvolvimento em TI e Formação (UFRGS, PPgECi e CNPq), Pesquisador e Membro do Projeto Manna Team (PCC/DIN/UEM/CNPq) e foi Professor e Pesquisador da Universidade de Brasília. Tem experiência na área de Sociologia, com ênfase em Sociologia da Ciência e da Tecnologia, da Educação e da Saúde; Educação em Saúde e Divulgação C&T; Desenvolvimento em C, T&I para TI e para Saúde; Gestão de C&T e Gestão de Pessoas no Serviço Público; Avaliação de Políticas Públicas principalmente Tecnologia de Informação e Comunicação (CC, ME, ID); e Cientometria.

E-mail: alexandre.motta@cnpq.br

Recebido em: 15/05/2023

Aprovado em: 15/10/2023