



Escola Nacional de Administração Pública

ESCOLA NACIONAL DE ADMINISTRAÇÃO PÚBLICA

**PROGRAMA DE MESTRADO PROFISSIONAL EM AVALIAÇÃO E
MONITORAMENTO DE POLÍTICAS PÚBLICAS**

**AVALIAÇÃO DO EFEITO DA BOLSA DE PRODUTIVIDADE
EM PESQUISA DO CNPQ NA CARREIRA DO PESQUISADOR
BRASILEIRO**

DISSERTAÇÃO DE MESTRADO

MARIA MARONY SOUSA FARIAS

BRASÍLIA – DF

2022

AVALIAÇÃO DO EFEITO DA BOLSA DE PRODUTIVIDADE EM PESQUISA DO CNPQ NA CARREIRA DO PESQUISADOR BRASILEIRO

Dissertação apresentado ao Programa de Avaliação e Monitoramento da Escola Nacional de Administração Pública - ENAP como requisito para obtenção do título de Mestre em Avaliação e Monitoramento de Políticas Públicas.

Aluna: Maria Marony Sousa Farias

Orientadora: Flávia Lúcia Chein Feres

Brasília - DF

2022

F224a Farias, Maria Marony Sousa

Avaliação do efeito da bolsa de produtividade em pesquisa do CNPq na carreira do pesquisador brasileiro / Maria Marony Sousa Farias. -- Brasília: Enap, 2022.

129 f. : il.

Dissertação (Mestrado -- Programa de Mestrado em Avaliação e Monitoramento) -- Escola Nacional de Administração Pública, 2022.

Orientação: Flávia Lúcia Chein Feres.

1. Produtividade. 2. Pesquisador. 3. Bolsa de Estudo. 4. Instituição de Pesquisa. I. Título. II. Feres, Flávia Lúcia Chein orient.

CDD 351.07

Bibliotecária: Tatiane de Oliveira Dias – CRB1/2230

MARIA MARONY SOUSA FARIAS

**AVALIAÇÃO DO EFEITO DA BOLSA DE PRODUTIVIDADE
EM PESQUISA DO CNPQ NA CARREIRA DO PESQUISADOR
BRASILEIRO**

Dissertação apresentado ao Programa de Mestrado em Avaliação e Monitoramento da Escola Nacional de Administração Pública - ENAP como requisito para obtenção do título de Mestre em Avaliação e Monitoramento de Políticas Públicas.

Defendida em 31 de Outubro de 2022.

BANCA EXAMINADORA:

Prof^a. Dra. Flávia Lúcia Chein Feres, Orientadora – ENAP

Prof. Dr. Vitor Azevedo Pereira Pontual – ENAP

Prof. Dr. Ivan Rocha Neto – Universidade Católica de Brasília - UCB

BRASÍLIA

2022

DEDICATÓRIA

A Deus e Nossa Senhora de Fátima, onde busco minha fortaleza.
A todo o curso de Mestrado em Avaliação e Monitoramento de Políticas Públicas da Escola Nacional de Administração Pública, corpo docente e discente, a quem fico lisonjeado por ter
feito parte.
Ao meu filho.
Ao meu marido e familiares pela compreensão.
Aos meus colegas de trabalho do CNPq.
Dedico este trabalho a todos os que me ajudaram ao longo desta caminhada.

AGRADECIMENTOS

Primeiramente a Deus, por todas as oportunidades e benção a mim concedidas.

A minha mãe e meus irmãos, por todo apoio em todas as situações de minha vida. Obrigada por acreditarem no meu potencial, por todos os conselhos, por todas as palavras de conforto em momentos difíceis, pela sabedoria a mim repassada, por todos os valores cultivados, por toda a motivação.

Ao meu filho e ao meu marido pelo apoio incondicional e pela compreensão das horas ausentes. Amo fortemente vocês.

Aos meus amigos no CNPq pelos comentários e sugestões.

Aos meus amigos de curso Nazareth, Patrício e Yuka pelos momentos compartilhados nas disciplinas e mesmo as nossas vidas tomando rumos diferentes, sempre estiveram presente.

Obrigada por cada instante compartilhado, por cada palavra amiga. Aprendi muito com vocês.

A minha orientadora Flávia Chein, pela atenção e suporte para realização deste trabalho, e a todos os demais professores do Mestrado em Avaliação e Monitoramento de Políticas Públicas, pelos ensinamentos e sempre se mostrarem dispostos a ajudar.

À ENAP pela oportunidade de fazer este curso.

Ao CNPq pelo suporte e sem o qual este trabalho não teria sido realizado.

LISTA DE SIGLAS E ABREVIATURAS

CA-AG - Comitê Assessor de Agronomia

CA-BF	- Comitê Assessor de Biofísica, Bioquímica, Farmacologia, Fisiologia e Neurociências
CA-CA	- Comitê Assessor de - Engenharia e Ciências Ambientais
CA-CS	- Comitê Assessor de Antropologia, Arqueologia, Ciência Política, Direito, Relações Internacionais e Sociologia
CA-EE	- Comitê Assessor de Engenharias Elétrica e Biomédica
CA-FA	- Comitê Assessor de Física e Astronomia
CA-LL	- Comitê Assessor de Letras e Linguística
CA-MD	- Comitê Assessor de Medicina
CA-MM	- Comitê Assessor de Engenharias de Minas e de Metalúrgica e Materiais
CA-VT	- Comitê Assessor de Medicina Veterinária
CA-ZO	- Comitê Assessor de Zoologia
CNPq	- Conselho Nacional de Desenvolvimento Científico e Tecnológico
DABS	- Diretoria de Ciências Agrárias, Biológicas e da Saúde
DEHS	- Diretoria de Engenharias, Ciências Exatas, Humanas e Sociais
PICC	- Plataforma Integrada Carlos Chagas
CGSAU	- Coordenação-Geral do Programa de Pesquisa em Saúde
CGCTM	- Coordenação-Geral do Programa de Pesquisa em Ciências da Terra e do Meio Ambiente-
CGAPB	- Coordenação-Geral do Programa de Pesquisa em Agropecuária e Biotecnologia
COCEX	- Coordenação do Programa de Pesquisa em Ciências Exatas
CGCEX	- Coordenação Geral de Programas de Pesquisa em Ciências Exatas
CGECT	- Coordenação Geral de Engenharia, Capacitação Tecnológica e Inovação
CGCHS	- Coordenação Geral do Programa de Pesquisa em Ciências Humanas e Sociais Aplicadas
	- Coordenação do Programa de Pesquisa em Biociências
COBIO	
COSAU	- Coordenação do Programa de Pesquisa em Saúde
COIAM	- Coordenação do Programa de Pesquisa Oceanográfica e Impactos Ambientais
COGEC	- Coordenação do Programa de Pesquisa em Gestão de Ecossistemas
COAGR	- Coordenação do Programa de Pesquisa em Agropecuária e do Agronegócio
COCQG	- Coordenação do Programa de Pesquisa em Ciências Químicas e Geociências
COCHS	- Coordenação do Programa de Pesquisa em Ciências Humanas e Sociais
COSAE	- Coordenação do Programa de Pesq. em Ciências Sociais Aplicadas e Educação
COENG	- Coordenação do Programa de Pesquisa em Engenharias
COENE	- Coordenação do Programa de Pesquisa em Energia
COSAO	- Coordenação de Suporte às Atividades Operacionais

COEIN	- Coordenação de Ecossistema de Informação
VIF	- Variance Inflation Factor
RH	- Recursos Humanos
JCR	- Journal Citation Reports
ISI	- International Scientific Indexing
CT&I	- Ciência, tecnologia e inovação
PCA	- Principal Component Analysis
ACP	- Análise de Componentes Principais
MQO	- Mínimos Quadrados Ordinários
TCC	- Trabalho de conclusão de curso

LISTA DE TABELAS

Tabela 5.1 – Teste-t de médias

LISTA DE FIGURAS

Figura 2.1 - Um breve histórico da institucionalização da avaliação por pares no CNPq.

Figura 3.1 – Descrição da Construção do Banco de dados obtido.

Figura 4.1 – Diagrama de dispersão para as variáveis do modelo. Elaboração própria a partir do Software R versão 4.1.2.

LISTA DE GRÁFICOS

Gráfico 2.1 – Distribuição das bolsas PQ, por gênero

Gráfico 2.2 – Distribuição das bolsas PQ, por categoria/nível

Gráfico 2.3 – Distribuição de bolsas PQ por Gênero e Comitê Assessor

Gráfico 2.4 – Distribuição da média de idade dos bolsistas PQ por comitê assessor e separadas por gênero

Gráfico 2.5 – Distribuição de bolsas PQ segundo a raça/cor declarada

Gráfico 2.6 – Distribuição de Bolsas PQ por Comitê Assessor

Gráfico 2.7 - Distribuição de bolsas PQ por Grandes Áreas do conhecimento

Gráfico 2.8 - Distribuição de bolsas PQ por região

Gráfico 2.9 - Distribuição de bolsas PQ por Estado

Gráfico 5.1 – Média da diferença entre o indicador de obtenção de capital científico (IND_CC) do grupo de controle e tratamento e seus respectivos limites inferiores e superiores dos intervalos de confiança (95%)

Gráfico 5.2 – Média do indicador de obtenção de capital científico (IND_CC) do grupo de controle e tratamento

LISTA DE QUADROS

Quadro 1.1 - Número de bolsas PQ e o respectivo investimento segundo Categoria/Nível

Quadro 2.1 - Distribuição dos bolsistas PQ nas grandes áreas, por gênero

Quadro 3.1 – Número de Propostas submetidas nas Chamadas de Bolsas PQ de 2013 a 2020.

Quadro 3.2 – Evolução dos Indicadores apresentados nas Planilhas de Produtividade de 2013 a 2020.

Quadro 3.3 – Comitês Analisados e suas respectivas diretorias, coordenações gerais e coordenações técnicas

Quadro 3.4 – Número de bolsas PQ dos comitês selecionados

Quadro 3.5 – Distribuição das variáveis nas planilhas de produtividade atualmente

Quadro 3.6 - Dimensão Esforço de Publicação - Planilha de Produtividade (2013, 2014, 2017 e 2020)

Quadro 3.7 - Dimensão Impacto das Publicações - Planilha de Produtividade (2013, 2014, 2017 e 2020)

Quadro 3.8 - Dimensão Esforço de Formação de Recursos Humanos - Planilha de Produtividade (2013, 2014, 2017 e 2020)

Quadro 3.9 - Dimensão Esforço de obtenção de capital científico - Planilha de Produtividade (2013, 2014, 2017 e 2020)

Quadro 3.10 – Percentual da variabilidade explicada pela componente 1 (PC1)

Quadro 3.11 – Número Total de Propostas submetidas nas Chamadas de Bolsas PQ em 2013, 2014, 2017 e 2020

Quadro 3.12 – Resultado do Painel obtido em número de indivíduos

Quadro 3.13 – Distribuição do Painel por gênero

Quadro 3.14 – Distribuição do Painel por região

Quadro 3.15 – Distribuição do Painel por Comitê

Quadro 4.1 – Método de Diferenças em Diferenças utilizando Grupo de Tratamento e Grupo de Controle

Quadro 4.2 – Descrição das variáveis explicativas do modelo proposto

Quadro 5.1 – Resultado da regressão para estimação do efeito de uma bolsa PQ na obtenção de capital científico (2020 em relação a 2013)

Quadro 5.2 – Resultado da regressão para estimação do efeito de uma bolsa PQ na obtenção de capital científico (2017 em relação a 2013)

RESUMO

As bolsas de produtividade PQ foram criadas na década de 1970, para subsidiar o trabalho do CNPq. Desde então, se mantêm como um instrumento de diferenciação simbólica entre pares e foi ganhando *status* ao longo do tempo e atualmente, se institui quase como uma carreira, formando um sistema hierarquizado de posições, no qual classifica e tipifica o perfil de excelência da elite científica brasileira. Esta dissertação buscou avaliar os efeitos da obtenção de uma bolsa de Produtividade PQ na carreira dos pesquisadores brasileiros. No estudo empírico foram analisados os dados de produtividade de pesquisadores de 12 comitês assessores referentes aos dois maiores comitês de cada uma das seis coordenações gerais do CNPq, os quais englobam todas as grandes áreas e cerca de 45% das bolsas PQ vigentes. Esses dados foram reunidos em um painel de indicadores de produtividade científica dos pesquisadores presentes na amostra referentes aos anos de 2013, 2014, 2017 e 2020. Para modelar e identificar o efeito de um conjunto de determinantes da obtenção de uma bolsa de produtividade PQ, foram estimados modelos de dados em painel com o estimador de diferenças em diferenças (DID). Os resultados obtidos apontam que a obtenção de uma bolsa PQ, aumenta em cerca de 5 unidades o indicador criado para refletir a carreira dos pesquisadores. No entanto, tal resultado só aparece quando podemos, de fato, isolar os indivíduos que não possuem bolsa PQ e compará-los com aqueles que possuem, o que foi possível para o ano de 2017, mas não de 2020. O estudo evidenciou que dentro o grupo de controle e do grupo de tratamento há bastante mobilidade, com pesquisadores que entram e saem da amostra. Os resultados também parecem indicar: 1) há, de fato, um enorme esforço por parte dos pesquisadores brasileiros não apenas para a obtenção destas bolsas de produtividade PQ, mas também para sua continuidade e para progressão de nível; e 2) o processo de julgamento efetuado pelos comitês assessores é de fato bem-feito, sendo necessário que os pesquisadores se mantenham continuamente produtivos e concorrendo com seus pares, caso queiram manter a bolsa ou obter progressão, caso contrário, irão perdê-la no ciclo seguinte.

Palavras Chaves: bolsa de produtividade PQ; capital científico, dados em painel, estimador de diferenças em diferenças.

ABSTRACT

The PQ productivity grants were created in the 1970s to subsidize the work of the CNPq. Since then they have remained an instrument of symbolic differentiation between peers and gained status over time. Today it is instituted almost as a career, forming a hierarchical system of positions, in which it classifies and typifies the profile of excellence of the Brazilian scientific elite. This dissertation sought to evaluate the effects of obtaining a PQ Productivity grant on the careers of Brazilian researchers. In the empirical study, the productivity data of researchers from 12 advisory committees referring to the two most significant committees of each of the six general coordination of CNPq, which encompass all major areas and about 45% of PQ grants, were analyzed. These data were gathered in a panel of scientific productivity indicators of the researchers present in the sample for 2013, 2014, 2017, and 2020. To model and identify the effect of a set of determinants of obtaining a PQ productivity grant, estimated panel data models with the difference-in-differences (DID) estimator. The results indicate that obtaining a PQ grant increases the indicator created to reflect the researchers' careers by about 4.5 units. However, such a result only appears when we can isolate individuals who do not have a PQ grant and compare them with those who do, which was possible for the year 2017, but not for 2020. The study showed that there is a lot of mobility within the control group and the treatment group, with researchers coming in and out of the sample. The results also seem to indicate two things: 1) there is, in fact, an enormous effort on the part of Brazilian researchers not only to obtain these PQ productivity grants, but also for their continuity and level progression; and 2) the judging process carried out by the advisory committees is in fact well done, requiring researchers to remain continuously productive and competing with their peers, if they want to keep the scholarship or level up or otherwise, they will lose it in the next cycle.

Keywords: PQ productivity grant; scientific capital, panel data, difference-in-differences estimator.

SUMÁRIO

LISTA DE SIGLAS E ABREVIATURAS.....	vii
LISTA DE TABELAS.....	x
Tabela 5.1 – Teste-t de médias	x
LISTA DE FIGURAS.....	x
LISTA DE GRÁFICOS	xii
LISTA DE QUADROS	xiii
RESUMO	xiv
ABSTRACT	xv
SUMÁRIO.....	xvi
1. INTRODUÇÃO	1
2. A AVALIAÇÃO POR PARES E SEU PROCESSO DE INSTITUCIONALIZAÇÃO NO CNPQ: UM BREVE HISTÓRICO SOBRE AS BOLSAS DE PRODUTIVIDADE PQ.....	4
2.1 As bolsas de produtividade no CNPq e seu papel para além da avaliação por pares.....	7
2.3 O perfil médio do pesquisador no Brasil: um panorama geral das bolsas PQ.....	11
3. BASE DE DADOS.....	19
3.1 Critério de Escolha dos Comitês.....	22
3.2 – Construindo o Painel de Dados	25
3.3 - Descrição das variáveis para a construção dos indicadores utilizados no modelo.....	29
3.3 Utilizando Análise de Componentes Principais para Redução de Dimensionalidade.....	35
3.4 Escolha dos Grupos de Tratamento e Controle	38
3.4.1 Caracterização dos Grupos de Tratamento e Controle	42
4. ESTRATÉGIA EMPÍRICA	44
4.1 Modelo de Regressão de Dados em Painel	44
4.1.1 Notação e tipos de regressão de dados em painel.....	45
4.1.2 Modelo MQO para dados empilhados (<i>pooled data</i>)	46
4.1.3 Modelo de Efeitos Fixos	47
4.1.4 Modelos de Efeitos Aleatórios	48
4.2 Modelo de Diferenças em Diferenças	49
4.3 Descrição das variáveis utilizadas no Modelo	51
4.4 Estatística descritiva das variáveis	54
5. RESULTADOS OBTIDOS.....	55
5.1 Análise dos Grupos de Tratamento e Controle	55
5.2 Resultados da Avaliação dos Efeitos da Bolsa de Produtividade PQ	59
5.3 Discussões sobre os resultados obtidos.....	64
6. CONSIDERAÇÕES FINAIS.....	67

7. REFERÊNCIAS BIBLIOGRÁFICAS	69
ANEXO A	73
ANEXO B	76
ANEXO C	79
APÊNDICE A	82
APÊNDICE B	87
APÊNDICE C	104

1. INTRODUÇÃO

As bolsas de produtividade foram criadas na década de 1970 e desde então, vem ocupando posição de destaque na hierarquia acadêmica brasileira. As bolsas de produtividade PQ foram criadas pelo CNPq como forma de incentivar os pesquisadores brasileiros a aumentarem sua produtividade em C&T e em contrapartida permitir que o CNPq tivesse um corpo consistente de consultores para apoiá-lo em suas decisões, na escolha dos projetos a serem apoiados ou não.

Desde sua criação, a bolsa PQ se mantém como um instrumento de diferenciação simbólica entre pares e com o tempo foi experimentando uma considerável alteração de status, evoluindo como um “título” e praticamente se instituindo como uma carreira e ideal a ser alcançado pelos pesquisadores no Brasil. Neste sentido, as bolsas de produtividade PQ são uma espécie de capital científico (conceito abordado no capítulo seguinte), pois além de ser um apoio financeiro, a bolsa de produtividade é um sinal de reconhecimento pelos pares aos pesquisadores que possuam uma destacada produção científica. Não obstante, a bolsa PQ permite a participação dos pesquisadores em uma série de processos fundamentais para a ciência brasileira, como concessão de bolsas e auxílios à pesquisa, certificação e avaliação de programas de pós-graduação na Capes, formulação de políticas públicas em ciência e tecnologia e inovação, composição de corpo editorial de periódicos científicos, bem como participação na elaboração dos próprios critérios de acesso à própria bolsa PQ (CNPQ, 2021b), (LEITE & ROCHA Neto, 2017). A concessão de uma bolsa PQ implica desta forma, na possibilidade de captação de recursos indiretos, pois além da remuneração do pesquisador, ela agrega o financiamento de outras bolsas, cujos efeitos se fazem e se refletem de forma direta na estruturação das equipes de pesquisa (GUEDES *et al*, 2015).

Como consultores ad hoc do CNPq, os bolsistas PQ emitem pareceres que subsidiam as decisões dos CA's, dos Comitês Temáticos (CT's) e dos servidores do CNPq com respeito à concessão de bolsas nas diversas modalidades apoiadas (Bolsas no País e exterior, bolsas de pesquisa e bolsas de fomento tecnológico), concessão de auxílios a projetos de pesquisa, dentre outros pareceres. Além de integrar o corpo de membros dos comitês assessores, um bolsista PQ pode passar a integrar a administração do CNPq, sendo um Diretor, membro do CD (Comitê Deliberativo) ou até mesmo Presidente.

Conforme analisa Guedes *et al* (2015), a intensificação do interesse e da demanda por essa bolsa está relacionada à valorização que lhe é conferida pela política de C&T nesse contexto, a qual é traduzida em reconhecimento e poder acadêmicos (GUEDES *et al*, 2015).

Segundo a Resolução Normativa RN 028/2015 (CNPQ, 2015a), a bolsa PQ é dividida hierarquicamente em 3 categorias: i) Sênior (PQ-SR), ii) PQ-1 (subdividida nos níveis 1A, 1B, 1C e 1D); e iii) PQ-2. Em 2003, o CNPq institui o chamado “Adicional de Bancada” (*Grant*), inicialmente apenas para os bolsistas PQ em nível 1A e 1B. Os recursos do Adicional de Bancada teriam como finalidade despesas de capital e custeio exclusivamente relacionadas ao projeto de pesquisa. Posteriormente, em 2004, tal benefício foi estendido a todos os pesquisadores categoria 1 (LEITE; ROCHA Neto, 2017). Os valores (em setembro/2022) para cada modalidade e bolsa PQ estão apresentados no Quadro 1.

A bolsa PQ-SR busca reconhecer a contribuição dos pesquisadores que são líderes em seu campo de atuação e destacados entre seus pares. Foi criada em 2005 como uma nova categoria de bolsa PQ e, a partir de 2011, tornou-se uma modalidade de bolsa independente, destinada a pesquisadores que já foram bolsistas PQ nos níveis 1A ou 1B por no mínimo 15 anos, consecutivos ou não, (CNPQ, 2015a).

No Quadro 1.1 é apresentada a quantidade total de bolsas PQ, por categoria/nível implementadas no ano de 2022, bem como o investimento total previsto em cada uma delas.

QUADRO 1.1 - Número de bolsas PQ e o respectivo investimento segundo Categoria/Nível¹

Categoria/Nível	Nº de bolsistas	% de bolsistas	Valor Unitário (R\$)	Adicional de Bancada (Unitário)	Total (mensal)	Total (em Milhões de Reais, anual)
1A	1168	8%	1.500,00	1.300,00	3.270.400,00	39.244.800,00
1B	1200	8%	1.400,00	1.100,00	3.000.000,00	36.000.000,00
1C	1423	9%	1.300,00	1.100,00	3.415.200,00	40.982.400,00
1D	2459	16%	1.200,00	1.000,00	5.409.800,00	64.917.600,00
2	8955	58%	1.100,00	0	9.850.500,00	118.206.000,00
SR	228	1%	1.500,00	0	342.000,00	4.104.000,00
TOTAL	15433	100%			25.287.900,00	303.454.800,00

Fonte: (<http://memoria2.cnpq.br/web/guest/bolsistas-vigentes>. Acesso em 20/08/2022). Elaboração: Autora.

Previsão no Ano de 2022. Valores previstos na RN-015/2013 - Tabela de Valores de Bolsas e Taxas no País. Disponível em: < http://memoria2.cnpq.br/view/-/journal_content/56_INSTANCE_0oED/10157/971393>. Acesso em 20/08/2022.

O presente trabalho tem o objetivo de avaliar os efeitos da obtenção de uma bolsa de Produtividade em Pesquisa (PQ) na carreira dos pesquisadores brasileiros. Dessa forma, para se alcançar o objetivo desse estudo, é aplicado o método econométrico de dados em painel com o estimador de diferenças em diferenças. No estudo empírico foram analisados os dados de produtividade de pesquisadores de 12 comitês assessores referentes aos dois maiores comitês de cada uma das seis coordenações gerais do CNPq, os quais englobam todas as grandes áreas e cerca de 45% das bolsas de produtividade vigentes no CNPq. Esses dados foram reunidos em um painel de indicadores de produtividade científica dos pesquisadores presentes na amostra referentes aos anos de 2013, 2014, 2017 e 2020.

A presente proposta de pesquisa está dividida, além dessa introdução, em cinco partes. Na primeira, dividida em três grandes tópicos, temos a apresentação do processo de institucionalização da avaliação por pares no CNPq e um breve histórico das bolsas de produtividade. Em seguida, são apresentados os conceitos de campo e capital científico de Bourdieu e concluindo a primeira parte, é feita uma análise do perfil médio atual destes bolsistas quanto a sua distribuição por comitê, região, estado, idade, gênero, raça, grande área e comitê. Na segunda parte desta dissertação, é apresentado como foi construído o banco de dados. Na terceira, é apresentado o processo de construção dos indicadores utilizados no modelo econométrico de dados em painel. Na quarta, temos a estratégia empírica utilizada para obtenção dos resultados. Na sequência, temos a apresentação e discussão dos resultados obtidos, e por último temos as considerações finais.

2. A AVALIAÇÃO POR PARES E SEU PROCESSO DE INSTITUCIONALIZAÇÃO NO CNPQ: UM BREVE HISTÓRICO SOBRE AS BOLSAS DE PRODUTIVIDADE PQ

A avaliação é parte integrante do processo de construção do conhecimento científico e surgiu com a própria ciência, há mais de três séculos, embora seus mecanismos, critérios, instrumentos e procedimentos venham experimentando modificações desde seu surgimento. O que tem permanecido constante ao longo de todo este tempo é a percepção de que apenas os próprios cientistas podem avaliar o trabalho de outros cientistas. Ou seja, somente os cientistas, por sua formação e experiência, seriam capazes de emitir uma opinião confiável e bem fundamentada e este processo tem sido, de forma geral, denominado de revisão por pares ou julgamento por pares (*peer review*), (DAVYT; VELHO, 2000).

A avaliação por pares praticada no CNPq é do tipo *single-blind* (simples cego), no qual os consultores conhecem os avaliados, mas não o contrário. Mas, nem sempre foi assim. Na tese de Anderson Cleiton Fernandes Leite é feito um apanhado sobre a institucionalização da avaliação por pares no CNPq (LEITE, 2018). Segundo Leite (2018), na criação do CNPq em 1951 não existia ainda uma normatização acerca das condições para a concessão de bolsas ou auxílios e a avaliação por pares nesta fase, baseava-se na informalidade, onde os próprios cientistas os quais também eram administradores no CNPq emitiam tais pareceres e quando necessário, recorriam a colegas (CASTRO; SOARES, 1986; LEITE, 2018).

À medida que o número de solicitações de bolsas e projetos foi crescendo, se fez necessária uma sistematização deste processo de avaliação e, em 1976, são criados os Comitês Assessores (CAs), por meio da Resolução Normativa 01/1976. O propósito é dar suporte especializado ao CNPq, além de “Propor a ação fomentadora do CNPq na sua respectiva área, inclusive com a apreciação, em calendário e métodos indicados pela Administração do CNPq, das solicitações de apoio à pesquisa e à formação de recursos humanos” (CNPQ, 1976). A Resolução Normativa 01/1976 se limita a especificar que a escolha dos membros deve ser feita dentre os “pesquisadores de reputação, que tenham posição de liderança em sua área e dela possuam conhecimento global” e tal escolha era competência exclusiva do Presidente do CNPq (CNPQ, 1976). Ainda não há até esta época, critérios formais de avaliação por pares.

Em 1985 foi criado o Ministério da Ciência e Tecnologia, ao qual o CNPq passou a ficar vinculado. Em 1988 surgem os Consultores Ad Hoc, com Resolução Normativa 12/88 (CNPQ, 1988). Com esta Resolução, o ato de avaliar os projetos apresentados passa por uma série de

determinações processuais e delimitações quanto a especificidade e clareza os procedimentos a serem adotados, (LEITE, 2017). Esta mesma Resolução garante o anonimato do avaliador, ou seja, uma avaliação *single-blind*. Mas, em contrapartida, o consultor ad hoc tem a obrigação de guardar sigilo quanto à materialidade do objeto avaliado (CNPQ, 1988).

Mesmo com o estabelecimento dos consultores ad hoc em 1988, somente com Resolução Normativa RN 004/98 é que se rompe com a exclusividade do presidente do CNPq na escolha dos integrantes dos CAs e tal função passa para o Conselho Deliberativo (CD)² com “base nos nomes sugeridos pela comunidade científica e tecnológica, segundo critérios estabelecidos pelo CD”. Segundo a mesma norma, a indicação de membros dos CAs “será feita a partir de nomes sugeridos pela comunidade científica e tecnológica nacional, segundo critérios e procedimentos a serem fixados no Regimento Interno do CNPq” (CNPQ, 1998). Ou seja, o órgão começa a constituir seu corpo de avaliadores com base em uma complexa rede de consultas à própria comunidade científica.

A partir de 2003, determinou-se que cada CA estabeleceria critérios para a avaliação dos proponentes à bolsa PQ. Os critérios deveriam seguir os seguintes parâmetros mínimos: publicação no site do CNPq, revisão a cada 03 anos e necessariamente apresentar aspectos quantitativos e qualitativos dentre seus itens (CNPQ, 2003b). Desde então, os Comitês de Assessoramento definem os critérios para se avaliar a produtividade dos pesquisadores em função da especificidade de cada área do conhecimento.

² O Conselho Deliberativo é a maior instância de poder decisório do CNPq, é formado pelo Presidente da Instituição, pelo secretário-executivo do Ministério da Ciência, Tecnologia e Inovações (MCTI) e representantes da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes/MEC), da Financiadora de Estudos e Projetos (FINEP), do Conselho Nacional das Fundações Estaduais de Amparo à Pesquisa (CONFAP), das comunidades científica, tecnológica e empresarial e dos servidores do CNPq. Disponível em: <<https://www.gov.br/cnpq/pt-br/composicao/conselho-deliberativo>>. Acesso em 16/11/2021.

Na Figura 2.1, temos um resumo do breve histórico apresentado.

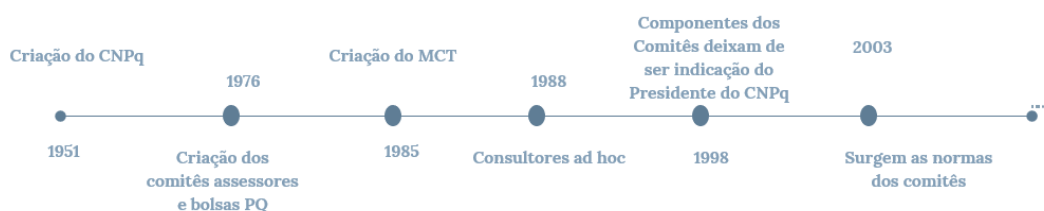


Figura 2.1 - Um breve histórico da sequência da institucionalização da avaliação por pares no CNPq.

Elaboração: Autora.

Ao se tornar bolsista PQ, o pesquisador passa a integrar a estrutura de avaliação por pares do CNPq, tornando-se obrigatoriamente um consultor *Ad Hoc*, (CNPQ, 2015b). Uma das funções básicas de bolsista de produtividade é a de atuar como consultor *ad hoc* no âmbito do CNPq e este trabalho é de fundamental importância para o funcionamento do órgão. Como consultores *ad hoc* do CNPq, os pesquisadores emitem pareceres que subsidiam as decisões dos CA's, dos Comitês Temáticos (CT's) e dos servidores do CNPq com respeito à concessão de bolsas nas diversas modalidades apoiadas (Bolsas no País e exterior, bolsas de pesquisa e bolsas de fomento tecnológico), concessão de auxílios a projetos de pesquisa, dentre outros pareceres.

Ressalte-se que, geralmente, o ingresso de novos bolsistas se dá na categoria 2. O perfil mínimo exigido para esta categoria é o de ser doutor há no mínimo 3 anos, sendo ele avaliado pela sua produção – trabalhos publicados e orientações – dos últimos 5 anos. Sua avaliação, classificação e progressão (ou rebaixamento) dentro das categorias e níveis da bolsa PQ é recomendada pelo Comitê ao qual está ligado e se dá pela submissão de um projeto de pesquisa em Chamada³ própria.

Além de demonstrar sua capacidade de formação de recursos humanos por meio de orientações de tese e dissertações, ao pretender atingir os níveis C e B, o pesquisador deve também acrescentar a seu currículo atividades relativas à organização de grupos de pesquisa e programas de graduação e pós-graduação. E no caso específico do nível B, espera-se que tenha papel ativo em órgãos de fomento à pesquisa. Por fim, o nível A é reservado a proponentes que continuamente tenham um elevado grau de excelência na produção científica, na formação de

³ A partir de 2016, a submissão de bolsas PQ passou a ser realizada em uma Chamada. Anteriormente, era realizada dentro de um calendário divulgado anualmente.

recursos humanos e na condução de grupos de pesquisa, demonstrando “uma significativa liderança dentro da sua área de pesquisa no Brasil e capacidade de explorar novas fronteiras científicas em projetos de risco” (CNPQ, 2015a).

2.1 As bolsas de produtividade no CNPq e seu papel para além da avaliação por pares

Desde sua criação, na década de 1970, a bolsa PQ se mantém como um instrumento de diferenciação simbólica entre pares e com o tempo houve uma alteração sensível de status, na qual evoluiu de um “título” na origem e se instituiu como uma carreira, formado por um sistema hierarquizado de posições, classificando e tipificando um perfil de excelência do que se pode denominar de uma elite científica (GUEDES *et al*, 2015).

Para compreender a questão de como uma bolsa de produtividade criada pelo CNPq em 1976 para valorizar e incentivar a produção científica no Brasil, acabou se tornando uma meta na carreira acadêmica brasileira, é necessário recorrer ao conceito de campo científico de Pierre Bourdieu.

O campo científico, é, conforme Bourdieu (1983), o lugar de uma luta concorrencial, onde o que está em jogo é o monopólio da autoridade científica, compreendida como a capacidade de falar e agir legitimamente (isto é, de maneira autorizada e com autoridade), que é socialmente outorgada a um agente determinado. Essa autoridade científica, pode ser traduzida como reputação, reputação que permite o acesso aos créditos e à influência em círculos mais amplos. A reputação vem com os resultados da pesquisa, do interesse que os pares têm por ela e da pertinência que lhe atribuem, em função de seus próprios trabalhos.

Assim, o trabalho de um cientista não deve ser importante apenas para ele, mas também para os outros, de tal forma que os cientistas tendem a se concentrar nos problemas considerados como os mais importantes, pois uma contribuição ou descoberta concernente a essas questões traz consigo um lucro simbólico mais importante (BOURDIEU, 1983).

Bourdieu (1983) define o campo científico como sendo:

“[...]um lugar de luta política pela dominação científica, que designa a cada pesquisador, em função da posição que ele ocupa, seus problemas, indissociavelmente políticos e científicos, e seus métodos, estratégias científicas que, pelo fato de se definirem expressa e objetivamente pela referência ao sistema de posições políticas e científicas constitutivas do campo científico, são ao mesmo tempo estratégias políticas. Não há “escolha” científica – do campo da pesquisa, dos métodos empregados, do lugar de publicação; ou ainda, escolha entre uma

publicação imediata de resultados parcialmente verificados e uma publicação tardia de resultados plenamente controlados – que não seja uma estratégia política de investimento objetivamente orientada para a maximização do lucro propriamente científico, isto é, a obtenção do reconhecimento dos pares-concorrentes” (BOURDIEAU, 1983, p.126-127).

Para Bourdieu (1983), a luta pela autoridade científica se constitui numa espécie particular de *capital social* que “...assegura um poder sobre os mecanismos constitutivos do campo e que pode ser reconvertido em outras espécies de capital...”. O capital aqui abordado por Bourdieu é o capital simbólico, aquele que se constitui por meio de uma relação social de conhecimento e reconhecimento entre os pares-concorrentes. Neste sentido o capital científico é um tipo de capital simbólico. É uma relação de produção e de reprodução de conhecimento e de reconhecimento especificamente dentro do campo científico.

Conforme Bourdieu (2004):

“Esse capital, de um tipo inteiramente particular, repousa, por sua vez, sobre o reconhecimento de uma competência que, para além dos efeitos que ela produz e em parte mediante esses efeitos, proporciona autoridade e contribui para definir não somente as regras do jogo, mas também suas regularidades, as leis segundo as quais vão se distribuir os lucros nesse jogo, as leis que fazem que seja ou não importante escrever sobre tal tema, o que é brilhante ou ultrapassado, e o que é mais compensador publicar no American Journal de tal e tal do que na Revue Française disso e daquilo”. (BOURDIEU, 2004, p.27).

O capital científico é, assim, um produto do reconhecimento, no qual Bourdieu (2004) distingue duas formas: uma dita “científica”, que repousa quase que exclusivamente sobre o reconhecimento pelos pares, pouco institucionalizada. Essa forma de capital científico está relacionada ao “prestígio”. Este primeiro tipo é o mais passível de ser contestado. A outra forma, dita “temporal” (ou política) remete ao poder político, poder institucional ou institucionalizado, que está ligado à ocupação de posições de poder nas instituições científicas, ao acesso aos meios de produção (créditos financeiros) e reprodução (participação em comissões, bancas de teses, bancas de concursos etc.) (BOURDIEAU, 2004; SHINN; RAGOUET, 2008).

As duas espécies de capital científico possuem leis de acumulação diferentes e segundo Bourdieu (2004), há duas formas de acumulação: a primeira é através de publicações, invenções ou descobertas (que seria o capital científico “puro”) e a segunda, seria através de estratégias políticas (capital científico institucionalizado), como participação em bancas de concursos, teses, dissertações, cerimônias, comissões etc.

Por tudo isso, pode-se afirmar que as bolsas de produtividade PQ, permitem aos pesquisadores a acumulação das duas formas de capital científico abordados por Bourdieu, pois além de ser um apoio financeiro, a bolsa de produtividade é um sinal de reconhecimento pelos pares aos pesquisadores que possuam uma destacada produção científica. Ao longo do seu percurso, o bolsista PQ pode tornar-se membro do CA (ou até mesmo adentrar à administração do CNPq, sendo um Diretor, membro do CD ou até mesmo Presidente. Ao perfazer esta trajetória, o pesquisador alcança o núcleo do campo científico no qual “[...] *os pesquisadores ou as pesquisas dominantes definem [...] o conjunto das questões que importam para os pesquisadores, sobre as quais eles vão concentrar seus esforços*” (BOURDIEU, 2004, p. 25).

No caso particular da pesquisa brasileira, a carreira científica “bem-sucedida” torna-se um processo contínuo de acumulação, no qual as bolsas PQ representariam a chancela dessa carreira acadêmica bem-sucedida. Dentro desse contexto, pode-se afirmar que as bolsas PQ são um termômetro da qualidade da pesquisa e um “selo” de qualidade do pesquisador e se tornaram uma meta na carreira da maioria dos cientistas brasileiros, e sendo assim é um elemento de diferenciação entre os pares. O processo de diferenciação no campo científico, segundo Cavalcante *et al* (2008), se distingue dos demais processos de diferenciação social, pela imposição de uma forma regulada e sistematizada de competição, pelos processos lógicos e experimentais a que estão sujeitas as formas de conhecimento e pela finalidade do conhecimento produzido. O processo de diferenciação que dá origem aos diversos campos científicos implica, portanto, a especificação de um objeto próprio e de agentes com formação específica para nele atuarem (CAVALCANTE, R. *et al.*, 2008).

A carreira científica é dedicada aos estudos e a produção de novos conhecimentos, tendo como base a pesquisa, o que requer um longo processo de formação, que muitas vezes, inicia-se bem antes da entrada formal no mundo acadêmico (AQUINO, 2006). Esse processo de formação contínua, vai avançando pela graduação, Mestrado e Doutorado e muitas vezes pós-doutorado. Trata-se de uma carreira muito longa e que leva muito tempo, antes de se alcançar uma posição de autoridade científica e acima de tudo, o reconhecimento dos pares. O futuro pesquisador brasileiro desde o Ensino Médio já tem plena consciência do papel da competição e do prestígio de seu futuro êxito. Sendo assim, esforça-se para obter as mais altas notas e passar em processos seletivos (como o Enem) para garantir um lugar nas melhores Universidades. Cursar uma graduação numa instituição respeitada e obter notas altas é decisivo para garantir a estima de seus professores e assim obter bolsas e cartas de recomendação. Para aqueles à procura de um emprego, vir de uma instituição reconhecida e ter trabalhado com algum

pesquisador renomado, pode garantir uma melhor posição e uma melhor avaliação. E esse processo continua com o acesso aos cargos administrativos, às comissões de governo, bancas de concursos, bancas de avaliação de mestrado, doutorado etc. Ou seja, o pesquisador depende de sua reputação junto aos pares para obter fundos para as suas pesquisas, para atrair estudantes de qualidade, para obtenção de consultorias, para obtenção de bolsas e outros convites. E poucas coisas no Brasil trazem tanta reputação à carreira de um pesquisador, quanto uma bolsa de produtividade.

As bolsas de produtividade PQ são assim, uma espécie de capital científico, que permite aos cientistas a participação em uma série de processos fundamentais para a ciência brasileira, como concessão de bolsas e auxílios à pesquisa, certificação e avaliação de programas de pós-graduação na Capes, formulação de políticas públicas em ciência e tecnologia e inovação, composição de corpo editorial de periódicos científicos, bem como participação na elaboração dos próprios critérios de acesso à própria bolsa PQ (CNPQ, 2020b), (LEITE; ROCHA Neto, 2017). A concessão de uma bolsa PQ implica, desse modo, a possibilidade de captação de recursos indiretos, pois além da remuneração do pesquisador, ela agrega o financiamento de outras bolsas, cujos efeitos se fazem se refletem de forma direta na estruturação das equipes de pesquisa (GUEDES *et al*, 2015).

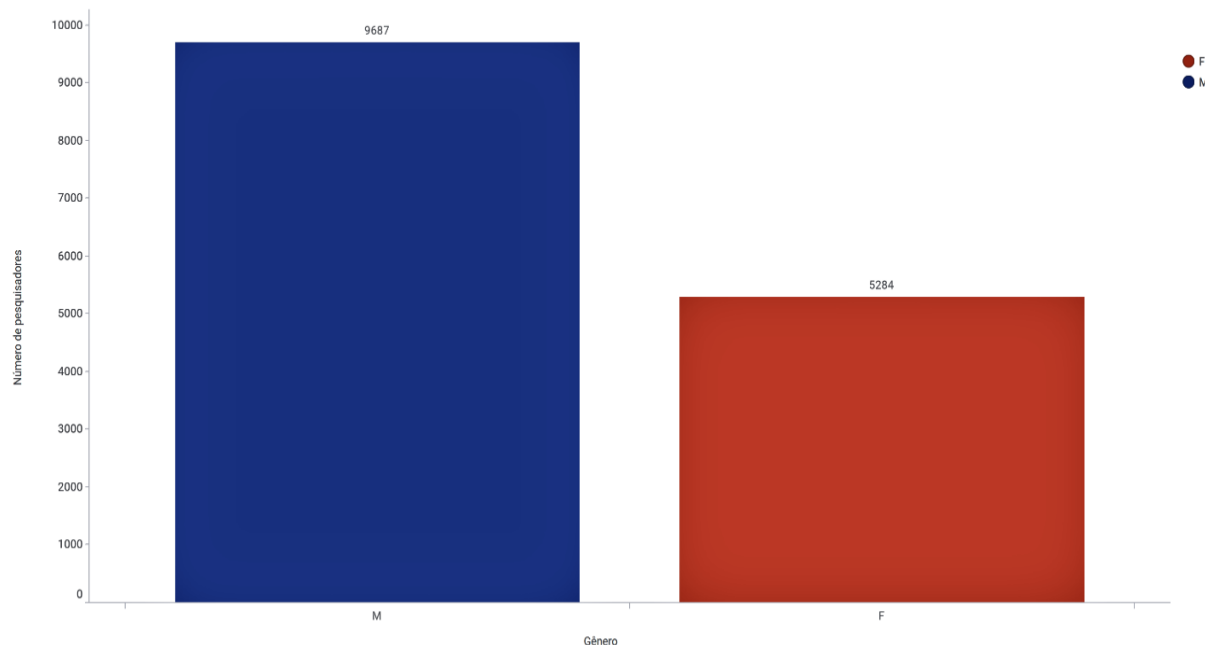
Por este motivo, a concessão de bolsas PQ tem sido objeto de intensa disputa e competitividade. Nesse contexto, entender em que medida a obtenção, manutenção ou perda de uma bolsa PQ afeta a carreira dos cientistas brasileiros é importante para entender a própria dinâmica da ciência no Brasil.

2.3 O perfil médio do pesquisador no Brasil: um panorama geral das bolsas PQ

Os dados apresentados neste tópico referem-se aos pesquisadores com bolsa PQ ativas em Janeiro de 2022. Os dados foram obtidos junto à COSAO - Coordenação de Suporte às atividades operacionais. O número total de bolsas PQ ativas em Janeiro de 2022 era de 14.971.

Um primeiro panorama a ser abordado é a distribuição por gênero das bolsas PQ no CNPq, apresentado no Gráfico 2.1. No Gráfico 2.1 pode-se observar uma distribuição desigual de bolsistas PQ por sexo, no qual temos que 65 % dos bolsistas PQ são homens e 35% são mulheres.

Gráfico 2.1 – Distribuição das bolsas PQ, por gênero

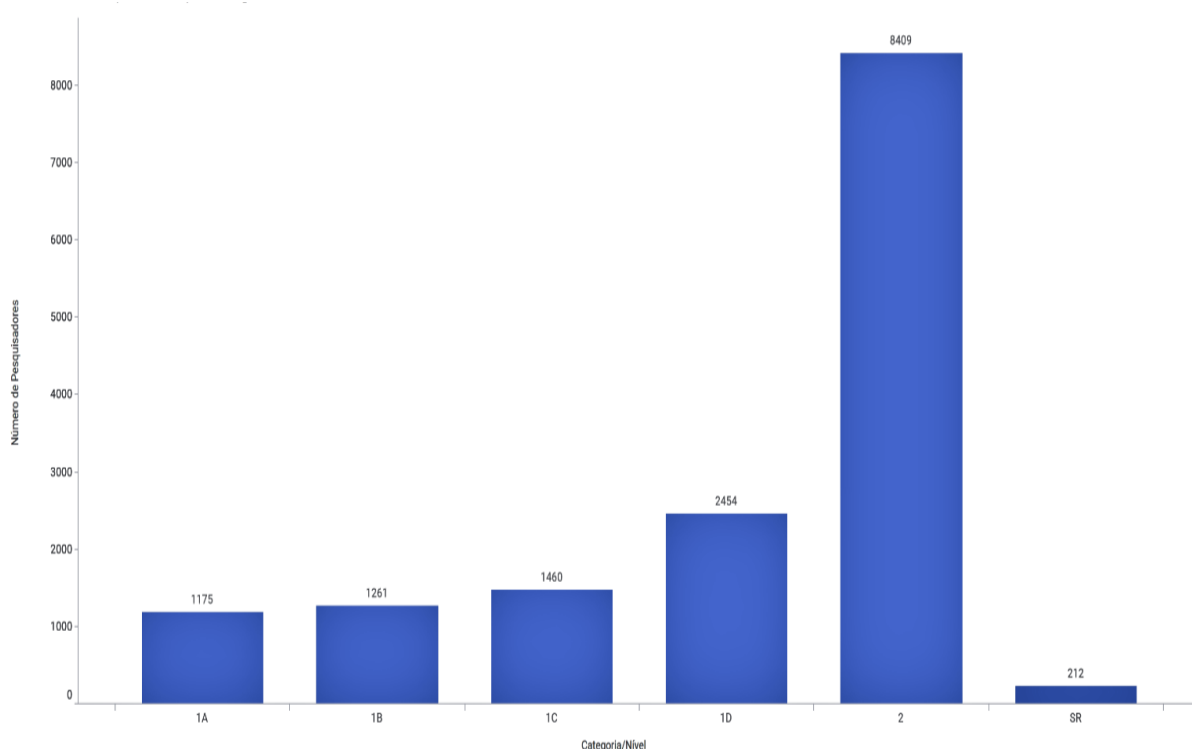


Nota: F – Feminino, M – Masculino. Fonte: COSAO/CNPq, 2022.

A distribuição das bolsas PQ, entre seus diferentes níveis, é apresentada no Gráfico 2.2. No Gráfico 2.2, temos os bolsistas PQ-2 representam 56 % dos casos; PQ-1D são 16 %; PQ-1C são 10 %; PQ-1B são 8 %; PQ-1A são 8 %; e PQ-Sr são 2%. Se considerarmos apenas a divisão existente entre o nível 2 e o nível 1 das bolsas PQ, temos que 63 % dos bolsistas nível 2 são homens e 37 % mulheres. Entre os bolsistas nível 1 temos 65 % de bolsistas homens e 35

% de bolsistas mulheres, o que nos aponta uma tendência de menor presença de mulheres nos níveis mais elevados desta modalidade de bolsa. Em todo o caso, esses dados parecem indicar que a distribuição desigual de bolsas PQ entre homens e mulheres é ligeiramente menor na categoria 2, onde estão as gerações mais recentes (majoritariamente com bolsistas nível 2).

Gráfico 2.2 – Distribuição das bolsas PQ, por categoria/nível



Fonte: COSAO/CNPq, 2022.

Ainda examinando a distribuição entre homens e mulheres nesses diferentes níveis de bolsa PQ, temos que no nível 1D os homens são 66 % dos bolsistas e as mulheres 34%; no nível 1C esses índices passam para 67 % e 33 %; no nível 1B são 68 % e 32 %; no nível 1A são 73% e 27 %; e no nível Sr são 74 % e 26%, para homens e mulheres, respectivamente. Esses números apontam para uma maior desigualdade nos níveis mais altos de bolsa PQ (PQ-1A e PQ-Sr).

Nesta pesquisa não cabe problematizar o porquê dessas desigualdades de gênero. No entanto, é interessante observar, que utilizando dados de 2002, Elisa Aquino (2004) apontava que as mulheres eram maioria entre os que receberam bolsas de Iniciação Científica e mestrado (respectivamente 55,1% e 52,9%) e quase metade entre os que receberam bolsas doutorado

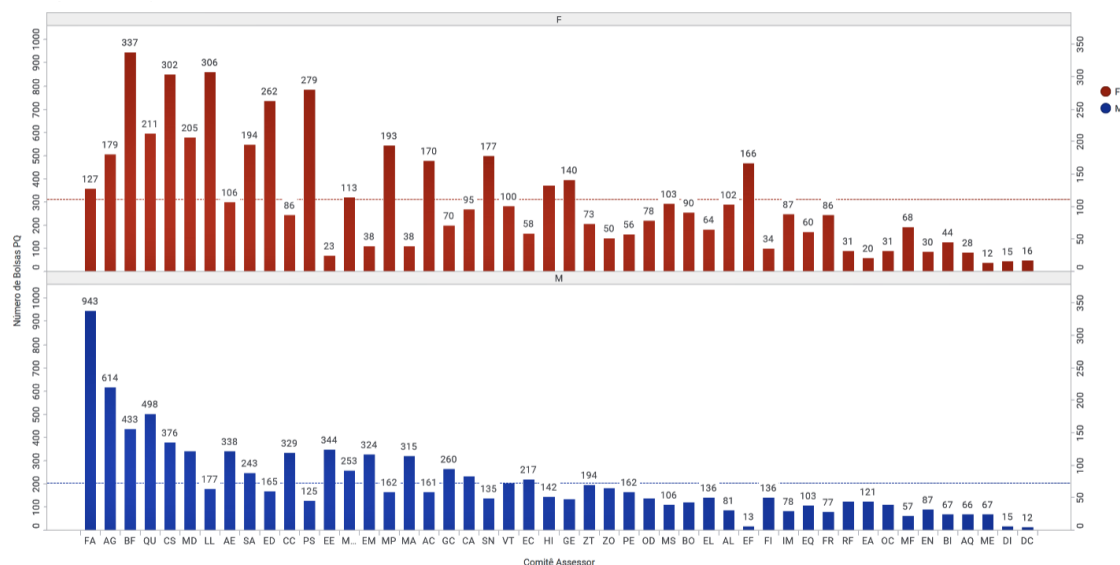
(48,4%) e de recém-doutor (49,7%) (Leta, 2003 *apud* Aquino, 2004). Porém, quando se trata de bolsas no exterior estes números são bem diferentes (36,5% em doutorado e 21,8% em pós-doutorado). Quanto as bolsas de produtividade, Elisa Aquino (2004), aponta que cerca de 67% são homens e que este gradiente aumenta à medida que aumenta o nível hierárquico da bolsa, (AQUINO, 2004).

Em 2006, Melo e Lastres (2006) analisaram a evolução das bolsas individuais do CNPq segundo modalidade e gênero, de 1990 a 1999 e constataram que os percentuais de mulheres com bolsa de pesquisa se mantiveram praticamente constantes durante toda a década de 1990, em torno de 30% e não apenas isso, o número de mulheres diminuía à medida que se aproximava dos níveis mais altos de bolsa de produtividade, (MELO; LASTRES, 2006).

Pode-se perceber que houve aumento na participação no número de mulheres com bolsas de produtividade PQ, mas muito pequeno se considerarmos o período de 3 décadas. Tais desequilíbrios identificados nesses estudos “resultam de um conjunto de obstáculos impostos às cientistas ao longo da carreira que dificultam o acesso ao topo do reconhecimento acadêmico” (ANDRADE, 2022). Um destes obstáculos é a maternidade. Conforme abordado por Andrade (2022), a maternidade ocorre no mesmo período em que as mulheres estão consolidando a carreira acadêmica e pesquisadoras que se tornam mães experimentam uma queda na produtividade acadêmica e isso pode ser um dos fatores que explicam esse desequilíbrio de gênero na pesquisa científica (ANDRADE, 2022). No entanto, como já mencionado, esse não é o objeto de estudo desta pesquisa, mas cumpre-nos mencionar o fato.

No Gráfico 2.3, temos a distribuição das bolsas PQ por gênero e Comitê Assessor. Pode-se observar que as mulheres são maioria nos comitês de Biofísica (BF) e Letras e Linguística (LL). Os comitês de Física e Astronomia (FA) e Agronomia (AG) possuem uma maior proporção masculina. Na Física, as mulheres possuem cerca de 12% do número total de bolsas PQ.

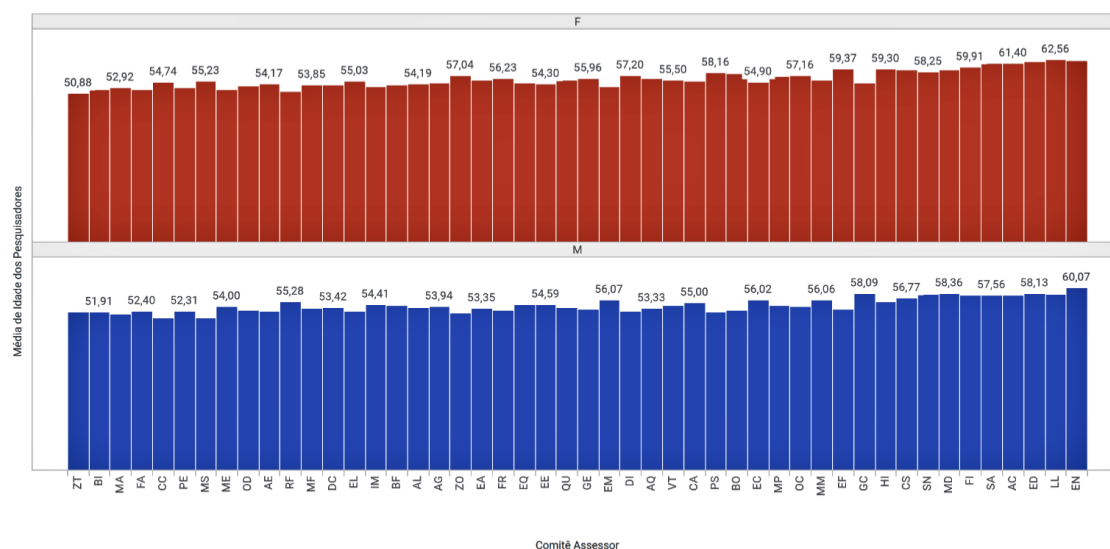
Gráfico 2.3 – Distribuição de bolsas PQ por Gênero e Comitê Assessor



Nota: F – Feminino, M – Masculino. Fonte: COSAO/CNPq, 2022.

No Gráfico 2.4, temos a média de idade dos bolsistas PQ, distribuídos por comitê assessor e separados por gênero. Na média, as mulheres têm 57 anos e os homens 54 anos. A média de idade é maior nos comitês de Engenharia Nuclear (EN) e Letras e Linguística (acima de 60 anos).

Gráfico 2.4 – Distribuição da média de idade dos bolsistas PQ por comitê assessor e separadas por gênero

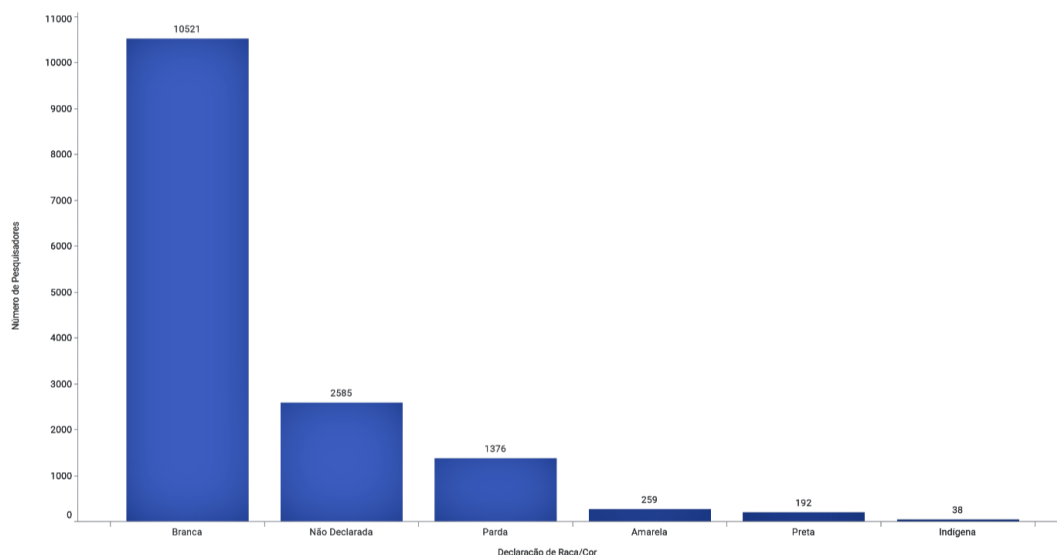


Nota: F – Feminino, M – Masculino. Fonte: COSAO/CNPq, 2022.

No Gráfico 2.5, temos a distribuição das bolsas PQ, por raça/cor. Dos 14.971 bolsistas em atividade, 70,3% se declaram como sendo da cor branca, 17,3% preferiram não declarar a

cor/raça, 9,2% se declararam como sendo da cor parda, 1,7% da cor amarela, 1,3% da cor preta e 0,2% como indígena. Do total, temos que 10% dos bolsistas se declaram como sendo da cor preta ou parda.

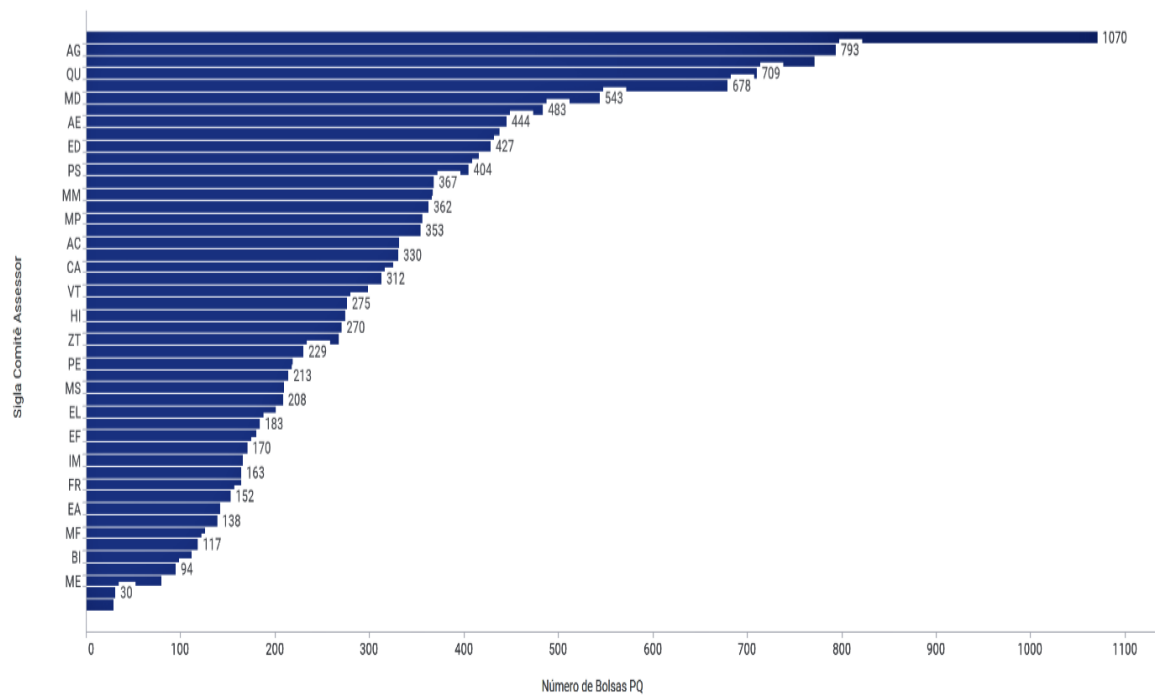
Gráfico 2.5 – Distribuição de bolsas PQ segundo a raça/cor declarada



Fonte: COSAO/CNPq, 2022.

No Gráfico 2.6, temos a distribuição das bolsas PQ, por Comitê Assessor. São 48 comitês, como já discutido anteriormente. O maior comitê do CNPq é o comitê assessor de Física e Astronomia (FA) com 1070 bolsistas PQ. Os 10 maiores comitês são formados pelos comitês assessores de Física e Astronomia (FA), Agronomia (AG), Biofísica (BF), Química (QU), Antropologia, Arqueologia, Ciência Política, Direito, Relações Internacionais e Sociologia (CS), Medicina (MD), Letras e Linguística (LL), Sociologia (CA), Administração e Educação (ED). Juntos, estes 10 comitês somam 42% de todas as bolsas PQ.

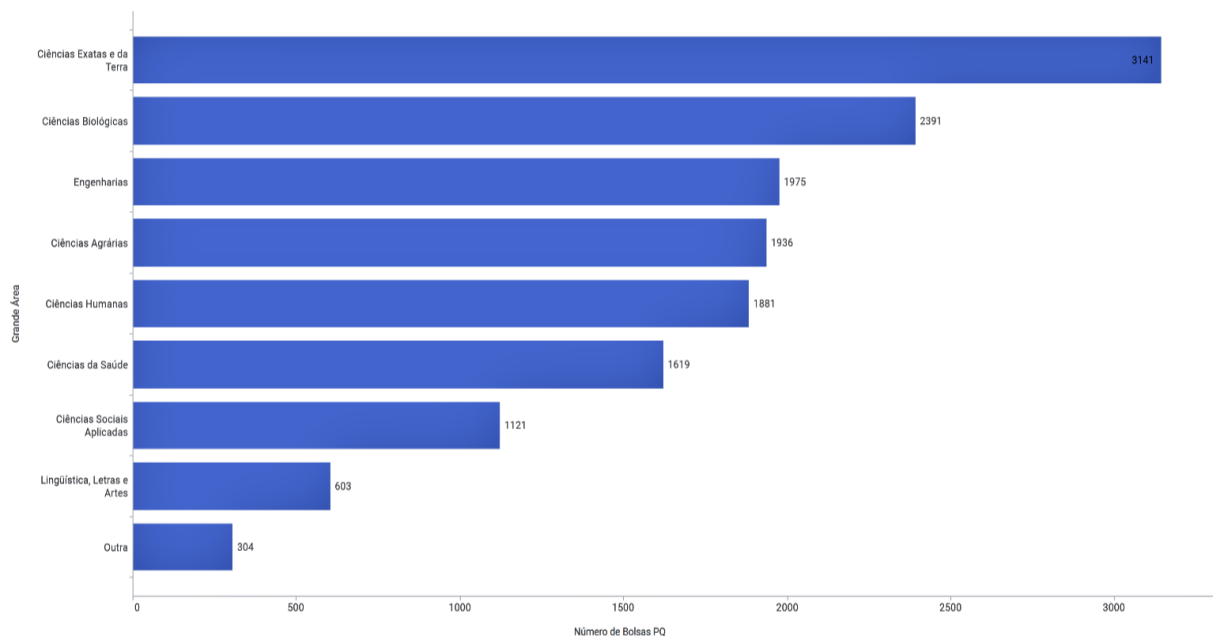
Gráfico 2.6 – Distribuição de Bolsas PQ por Comitê Assessor



Fonte: COSAO/CNPq, 2022.

No Gráfico 2.7, temos a distribuição de bolsas PQ por Grandes Áreas do conhecimento.

Gráfico 2.7 - Distribuição de bolsas PQ por Grandes Áreas do conhecimento



Fonte: COSAO/CNPq, 2022.

Examinando-se a distribuição com relação às diferentes grandes áreas do conhecimento, temos que as bolsas PQ estão distribuídas da seguinte forma entre as grandes áreas: Ciências

Exatas e da Terra 21%; Ciências Biológicas 16 %; Ciências Agrárias 13%, Ciências Humanas 12,5%, Ciências da Saúde 11%; Ciências Sociais Aplicadas 7,5%; Letras, Linguística e Artes 4%, Engenharias 13% e Outras 2%.

Com respeito à distribuição interna das Grandes Áreas entre homens e mulheres, temos:

Quadro 2.1 - Distribuição dos bolsistas PQ nas grandes áreas, por gênero

Grande Área	% Homens	%Mulheres
Ciências Exatas e da Terra	81	19
Ciências Humanas	50,1	49,9
Ciências Biológicas	56	44
Ciências da Saúde	49	51
Ciências Sociais Aplicadas	58	42
Ciências Agrárias	72	28
Letras, Linguística e Artes	38	62
Engenharias	79	21
Outras	70	30

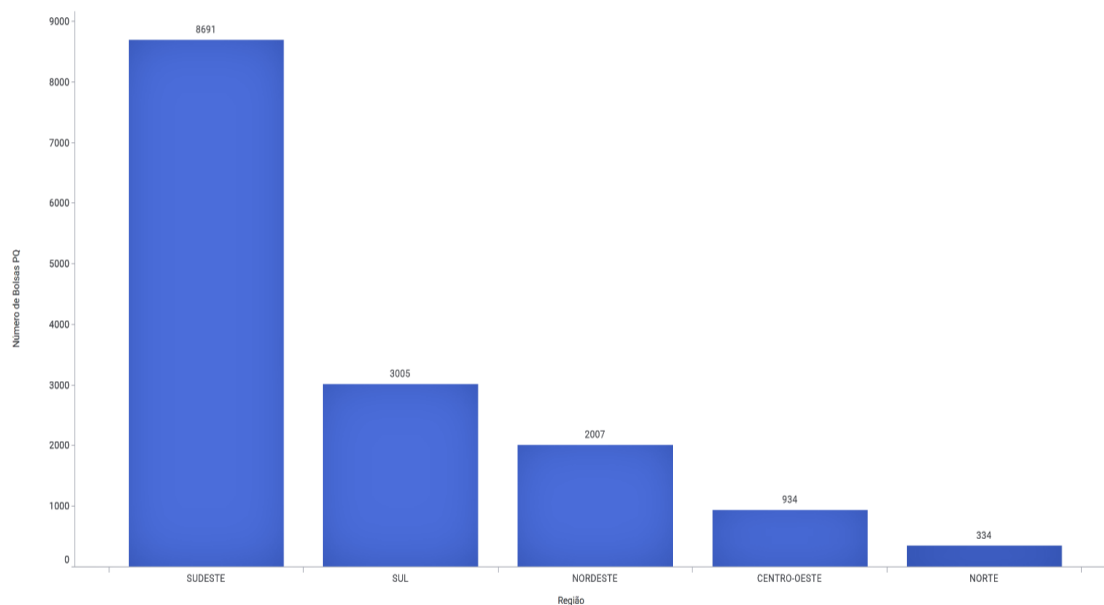
Fonte: COSAO/CNPq, 2022. Elaboração própria.

Pode-se observar no Quadro 2.1 que as mulheres são maioria apenas na Grande área de Letras, Linguística e Artes, onde são 62% do total de bolsas. Já na área de Ciências Exatas e da Terra e Engenharias, o percentual está em torno de 20% e nota-se certo equilíbrio nas Ciências Biológicas e Ciências da Saúde.

A distribuição de bolsas PQ por região está apresentada no Gráfico 2.8. Pode-se observar que a região Sudeste concentra 58% de todas as bolsas PQ, seguida pela Sul com 20%. A região Nordeste possui cerca de 13%, a região Centro-oeste 6% e a região Norte apenas 2%. Ou seja, a região Sudeste, possui mais bolsistas PQ que todas as outras regiões somadas e juntas, as regiões Sul e Sudeste concentram 78% das bolsas.

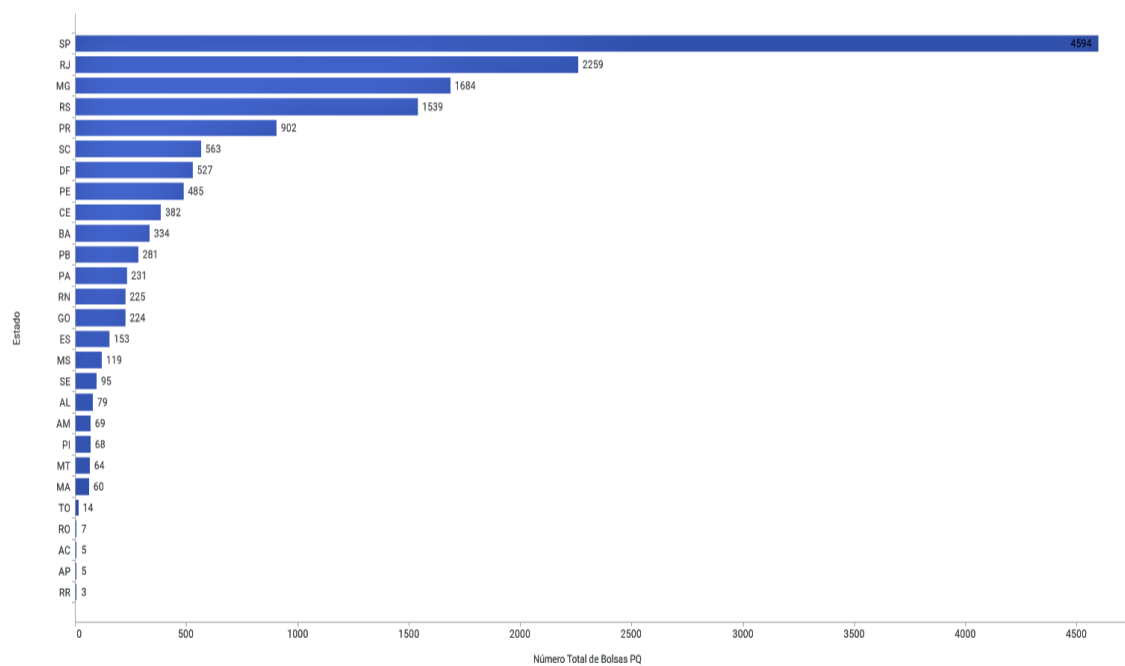
A distribuição de bolsas PQ por Estado está apresentada no Gráfico 2.9. No Gráfico 2.9, pode-se notar que os estados de São Paulo, Rio de Janeiro, Minas Gerais e Rio Grande do Sul detêm a maior parte das bolsas PQ, sendo que o estado de São Paulo sozinho possui 4594 bolsas (cerca de 31%). Juntos, os quatro estados citados anteriormente detêm cerca de 68% de todas as bolsas PQ. Sendo os estados do norte brasileiro aqueles com o menor número de bolsas.

Gráfico 2.8 - Distribuição de bolsas PQ por região



Fonte: COSAO/CNPq, 2022.

Gráfico 2.9 - Distribuição de bolsas PQ por Estado



Fonte: COSAO/CNPq, 2022.

3. BASE DE DADOS

A base de dados utilizada neste trabalho foi construída a partir da combinação de duas diferentes fontes: 1) as planilhas de deliberação final (CGEAO⁴/CNPq); e 3) as planilhas de indicadores bibliométricos (ou planilhas de produtividade, como são chamadas) fornecidas (COEIN⁵/CNPq), os quais são extraídos da base Lattes do CNPq.

As planilhas de deliberação final, correspondem as planilhas com o resultado final do julgamento das chamadas (deliberação da Diretorias DABS e DEHS) e informações dos projetos submetidos, tais como proponente, instituição, comitê etc. Em média, esta planilha contém cerca de dez mil (ou mais) solicitações de bolsa PQ, dependendo do ano, vide Quadro 3.1.

Quadro 3.1 – Número de Propostas submetidas nas Chamadas de Bolsas PQ de 2013 a 2020.

ANO	Total de propostas submetidas
2013	9766
2014	8974
2015	10199
2016	9434
2017	10199
2018	10976
2019	11596
2020	10138

Elaboração: Autora. Fonte: CGEAO/CNPq.

Analisar os dados bibliométricos de pesquisadores de um ano específico não é uma tarefa trivial, tendo em vista que estes atualizam constantemente seus dados na Base Lattes. Por este motivo, neste trabalho são utilizados os dados bibliométricos dos pesquisadores disponibilizados pelo próprio CNPq aos membros dos comitês assessores, por ocasião do julgamento.

Durante o processo de julgamento de bolsas PQ, o CNPq por meio da COEIN (Coordenação de Ecossistemas de Informação) disponibiliza para os membros dos comitês

⁴ CGEAO - Coordenação Geral de Apoio Operacional

⁵ COEIN – Coordenação de Ecossistema de Informação

assessores, planilhas com os dados bibliométricos de todos os pesquisadores que submeteram propostas, com o objetivo de auxiliar o trabalho destes comitês. Estes dados são extraídos da Base Lattes e refletem o currículo Lattes congelado dos pesquisadores concorrentes dos últimos 10 e 5 anos, dependendo da categoria/nível do pesquisador. São os dados destas planilhas que estão utilizados neste trabalho. Estas são as planilhas de indicadores, ou planilhas de produtividade, como são comumente chamadas pelos comitês.

Cada Comitê recebe uma planilha, contendo os dados dos pesquisadores que estão pleiteando bolsa naquele ano específico e cada Comitê filtra o que lhes é pertinente, dadas as especificidades de cada um. Essa planilha de produtividade, extraída da base Lattes, possui atualmente 101 indicadores sobre a produtividade dos pesquisadores, dentre os quais estão: número total de artigos, número total de patentes, número de publicações em revistas Qualis A, B e C, índice H, índice JCR, número de orientações de doutorado concluídas e em andamento, número de dissertações concluídas e em andamento, participação em grupos de pesquisa, participação em comitês editoriais, coordenação de projetos de pesquisa, número de livros publicados, número de citações, dentre outros.

Estas planilhas, que começaram a ser produzidas em 2013, nem sempre tiveram esta quantidade de indicadores. O número de indicadores foi evoluindo ao longo do tempo e de acordo com a necessidade dos comitês. No Quadro 3.2, temos de forma resumidas, as principais informações disponibilizadas nestas planilhas ao longo dos anos. Todas as variáveis utilizadas desde 2013 até os dias atuais estão descritas nos Anexos A, B e C desta dissertação.

Quadro 3.2 – Evolução dos Indicadores apresentados nas Planilhas de Produtividade de 2013 a 2020.

Informações	2013	2014	2015	2016	2017	2018	2019	2020
Número de Artigos Total e Número de artigos com JCR	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Livros, capítulos de Livros e Organização de obras	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Prêmios e Títulos	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Trabalhos publicados em Eventos	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Patentes	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Patentes Total	Não	Não	Não	Sim	Sim	Sim	Sim	Sim
Citações	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Coorientações de Mestrado e Doutorado	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Qualis A, B e C	Não	Não	Não	Sim	Sim	Sim	Sim	Sim
Coordenação de projetos de Pesquisa, Desenvolvimento ou Extensão	Não	Não	Não	Sim	Sim	Sim	Sim	Sim
Participação como membro em projetos de Pesquisa, Desenvolvimento ou Extensão	Não	Não	Não	Sim	Sim	Sim	Sim	Sim
Orientação de TCC	Não	Não	Não	Sim	Sim	Sim	Sim	Sim
Participação em Bancas de Mestrado e Doutorado	Não	Não	Não	Sim	Sim	Sim	Sim	Sim

Elaboração: Autora. Fonte: COEIN/CNPq.

3.1 Critério de Escolha dos Comitês

A princípio seriam analisados todos os comitês do CNPq, ou seja, 48 comitês. No entanto, dado o grande número de dados a serem tratados, essa ideia teve que ser descartada e uma nova estratégia foi abordada. Conforme já tratado neste documento, o CNPq possui duas diretorias ligadas ao fomento de Ciência, tecnologia e Inovação em todas as áreas do conhecimento: DABS e DEHS. Cada uma delas possui 24 comitês e 03 Coordenações Gerais. Desta forma, o critério estabelecido foi o de escolher os dois maiores comitês de cada Coordenação Geral, totalizando assim, 12 comitês. São eles:

- 1) CA-AG - Comitê Assessor de Agronomia;
- 2) CA-BF - Comitê Assessor de Biofísica, Bioquímica, Farmacologia, Fisiologia e Neurociências
- 3) CA-CA - Comitê Assessor de - Engenharia e Ciências Ambientais
- 4) CA-CS - Comitê Assessor de Antropologia, Arqueologia, Ciência Política, Direito, Relações Internacionais e Sociologia
- 5) CA-EE - Comitê Assessor de Engenharias Elétrica e Biomédica
- 6) CA-FA - Comitê Assessor de Física e Astronomia
- 7) CA-LL - Comitê Assessor de Letras e Linguística
- 8) CA-MD - Comitê Assessor de Medicina
- 9) CA-MM - Comitê Assessor de Engenharias de Minas e de Metalúrgica e Materiais
- 10) CA-QU - Comitê Assessor de Química
- 11) CA-VT - Comitê Assessor de Medicina Veterinária
- 12) CA-ZO - Comitê Assessor de Zoologia

As Diretorias, as Coordenações Gerais e as Coordenações Técnicas de cada comitê estudado estão apresentadas no Quadro 3.3.

Quadro 3.3 – Comitês Analisados e suas respectivas diretorias, coordenações gerais e coordenações técnicas

DIRETORIA	COORDENAÇÃO GERAL	COORDENAÇÃO TÉCNICA	COMITÊ ASSESSOR
DABS	CGSAU	COBIO	BF - Biofísica, Bioquímica, Farmacologia, Fisiologia e Neurociências
		COSAU	MD - Medicina
	CGCTM	COIAM	CA - Engenharia e Ciências Ambientais
		COGEC	ZO – Zoologia
	CGAPB	COAGR	AG - Agronomia
		COAGR	VT - Medicina Veterinária
DEHS	CGCEX	COCQG	QU - Química
		COCEX	FA - Física e Astronomia
	CGCHS	COCHS	LL - Letras e Linguística
		COSAE	CS - Antropologia, Arqueologia, Ciência Política, Direito, Relações Internacionais e Sociologia
	CGECT	COENG	EE - Engenharias Elétrica e Biomédica
		COENE	MM - Engenharias de Minas e de Metalúrgica e Materiais

Fonte: (CNPq1, 2021). Elaboração: Autora.

O CNPq possui atualmente 15423⁶ bolsas de produtividade PQ, conforme distribuição por comitê apresentada no Quadro 3.4. Pode-se notar no Quadro 3.4 que os comitês selecionados reúnem cerca de 44 % de todas as bolsas de produtividade do CNPq e, além disso, cobrem todas as grandes áreas do conhecimento (Ciências Exatas e da Terra, Ciências Biológicas, Ciências Agrárias, Ciências Humanas, Ciências Sociais Aplicadas, Ciências da Saúde, Letras, Linguística e Artes, Engenharias e Outras), constituindo assim uma amostra factível e razoável.

⁶ Este número de pesquisadores refere-se ao número de bolsistas PQ com bolsas ativas em Junho de 2022. Os números apresentados no Quadro 2.2 foram obtidos em Janeiro de 2022, quando havia 14971 bolsistas ativos.

Quadro 3.4 – Número de bolsas PQ dos comitês selecionados

Comitê Assessor - CA	Grande Área	Nº de bolsas PQ no CA	% em relação nº total de bolsas
AG	Ciências Agrárias	793	5,3%
BF	Ciências Biológicas e Outras	770	5,1%
CA	Ciências Exatas e da Terra, Engenharias e Outras	324	2,2%
CS	Ciências Humanas e Ciências Sociais Aplicadas	678	4,5%
EE	Engenharias e Outras	367	2,5%
FA	Ciências Exatas e da Terra	1070	7,1%
LL	Linguística, Letras e Artes	483	3,2%
MD	Ciências da Saúde	543	3,6%
MM	Engenharias	366	2,4%
QU	Ciências Exatas e da Terra	709	4,7%
VT	Ciências Agrárias	298	2,0%
ZO	Ciências Biológicas	229	1,5%
TOTAL		6630	44,3%

Fonte: Disponível em: (<http://memoria2.cnpq.br/bolsistas-vigentes>, acesso em Janeiro/2022). Elaboração: Autora.

3.2 – Construindo o Painel de Dados

Após selecionar os comitês a serem estudados, foi decidido qual período seria analisado, de forma a não desperdiçar tempo limpando e organizando planilhas desnecessariamente. A linha base de raciocínio parte da planilha de indicadores, pois o objetivo do trabalho é avaliar o efeito da bolsa de produtividade na carreira dos pesquisadores brasileiros utilizando dados em painel, combinado com método de diferenças em diferenças. A princípio seriam utilizados os dados dos anos de 2010 a 2021. No entanto, as planilhas de produtividade só começaram a ser produzidas para todos os comitês em 2013⁷, quando passaram a ser institucionalmente utilizadas. Assim, o ano de partida é o ano de 2013. Como uma bolsa PQ-2 tem duração de 36 meses, o período avaliado se dá a cada 3 anos.

A ideia central nesta pesquisa é acompanhar a carreira de pesquisadores ao longo do tempo, partindo de um período em que nenhum dos pesquisadores selecionados possui bolsa PQ. Desta forma, é possível acompanhar o desenvolvimento da carreira dos pesquisadores em diversos momentos: 1) um período anterior quando o pesquisador não possui bolsa (2013); 2) um segundo período em que o pesquisador consegue uma bolsa PQ-2 (2014); e 3) um terceiro e quarto períodos em que este mesmo pesquisador submete novamente a solicitação de bolsa (2017 e 2020).

Tendo em vista o comportamento dos pesquisadores, os quais costumam solicitar esta bolsa de forma recorrente (ano após ano, até conseguirem), os dados dos pesquisadores selecionados em 2013 são cruzados com os dados dos mesmos pesquisadores em 2014, 2017 e 2020. Assim, em 2014, são selecionados todos os pesquisadores, que obtiveram bolsa PQ-2. Aqueles que obtiveram bolsa vão constituir o Grupo de Tratamento e os que não conseguiram, vão constituir o Grupo de Controle. Essa é uma forma de captar os fatores que variam ao longo do tempo, medindo a mudança antes e depois dos resultados para o grupo que obteve bolsa e para o outro que não obteve, mas está exposto ao mesmo conjunto de condições ambientais.

Quem obteve bolsa em 2014, provavelmente vai submeter novo pedido de bolsa em 2017 e em 2020. Aqueles que não obtiveram, provavelmente, vão continuar tentando ao longo do período analisado. Assim, os mesmos pesquisadores selecionados em 2013, terão suas carreiras analisadas ao longo do período 2013-2020.

⁷ Antes disso, haviam sido produzidas de forma isolada para os comitês de Física e Química, por iniciativa de servidores.

Decidido qual período e quais os comitês a serem avaliados, os dados foram selecionados e tratados inicialmente ano a ano.

Nas planilhas de deliberação final foram selecionados apenas os dados dos comitês a serem estudados e excluídos os demais para diminuir o tamanho da planilha. Além disso, foram retiradas todas as informações que não seriam utilizadas. Foram mantidas apenas as informações relativas ao número de processo, nome do comitê, parecer de recomendação, parecer de aprovação, coordenação técnica, coordenação geral, diretoria, estado, região e instituição.

Além das informações constantes nas planilhas de deliberação final, foram acrescentadas variáveis que correspondem ao ano do julgamento da chamada, gênero (masculino ou feminino) e grupo (controle ou tratamento). Nenhuma das planilhas obtidas incluía a informação de gênero e tal informação teve que ser levantada manualmente por meio da Plataforma Lattes. Foram selecionadas apenas as planilhas referentes aos 12 comitês estudados, para os anos de 2013, 2014, 2017 e 2020. As informações desnecessárias foram retiradas e mantidas apenas aquelas referentes à produtividade e número de processo. Além das informações constantes nas planilhas de produtividade, foram acrescentadas variáveis que correspondem ao ano do julgamento da chamada, tempo de conclusão de doutorado (calculado a partir do ano de conclusão do doutorado em relação ao ano da chamada) e comitê (comitê no qual o pesquisador submeteu a solicitação de bolsa).

Todas as variáveis utilizadas, em todas as planilhas tiveram os nomes adaptados para o mesmo padrão das planilhas de produtividade, que é o de letras maiúsculas, nenhum acento e uso de símbolo () *underline*, no lugar de espaço. Isso facilita a documentação e a programação no software R. Neste trabalho foi utilizada a versão 4.1.2 do R.

Todo o trabalho descrito até aqui, foi para reunir as planilhas dos comitês em análise separadas por ano. Ou seja, ainda havia 8 planilhas, 2 para cada ano. Utilizando o Software R, as 2 planilhas (deliberação final e produtividade) referentes a cada ano analisado (2013, 2014, 2017 e 2020) foram reunidas em uma única planilha por meio do número de processo, que é único para pesquisador dentro de uma chamada. Desta forma, para cada ano foi criada uma única planilha com todos os dados e variáveis das planilhas mencionadas e o número de planilhas foi reduzido a 4.

Após esta etapa, na planilha referente aos dados de 2013, foram filtrados apenas os pesquisadores que obtiveram parecer de recomendação dos comitês, mas não obtiveram parecer de aprovação das Diretorias (não obtiveram bolsa). O objetivo disso é que todos os pesquisadores neste estudo comecem no mesmo ponto, ou seja, nenhum dos pesquisadores

havia sido contemplado com bolsa PQ no ano anterior à aplicação da política. A política aqui refere-se à obtenção de bolsa.

Finalmente, por meio do “cruzamento” de dados, utilizando a variável “CPF” como chave, todas as planilhas resultantes de cada ano foram reunidas utilizando o Software R, por meio da função *inner_join*. Ao final desta etapa as planilhas iniciais tornaram-se uma única planilha resultante. O banco de dados resultante, conta com 133 variáveis, descritas no Anexo II.

O processo de obtenção da base de dados final é mostrado na Figura 3.1.

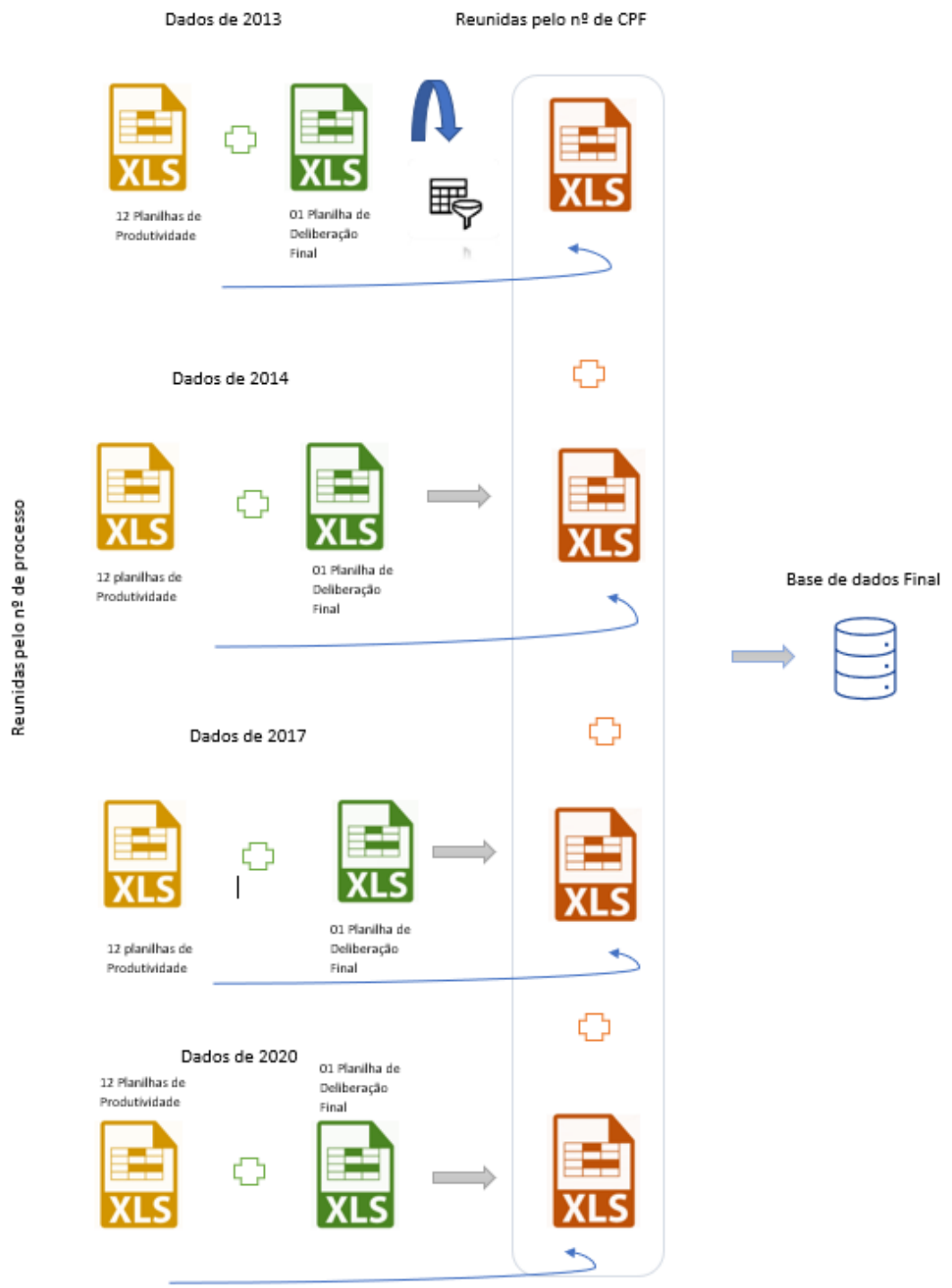


Figura 3.1 – Descrição da Construção do Banco de dados obtido. Fonte: elaboração própria.

A seguir, é feita uma descrição das características de Banco de dados obtido.

3.3 - Descrição das variáveis para a construção dos indicadores utilizados no modelo

A planilha de produtividade atual possui 39 variáveis que quantificam as publicações dos pesquisadores, incluindo artigos, livros, capítulos de livros, organização de obras e trabalhos em eventos; 19 variáveis que quantificam o impacto de tais publicações, 21 que quantificam a participação dos pesquisadores na formação de recursos humanos (orientação e coorientação de Trabalho de conclusão de curso (TCC), Iniciação científica, mestrado, doutorado, supervisão de pós-doutorado), 09 variáveis que quantificam a participação do pesquisador em projetos de pesquisa, desenvolvimento e extensão (seja como coordenador ou como membro de equipe), além de outras 08 variáveis que quantificam número de prêmios e títulos, número de participações em comitês assessores, em organizações de eventos e participação em bancas de mestrado e doutorado, além de mais 12 variáveis que contabilizam se o pesquisador é líder de grupos de pesquisa, se possui pós-doutorado prévio, patentes etc.

No Quadro 2.5, temos um resumo de como são distribuídas as 108 variáveis utilizadas na planilha de produtividade.

Quadro 3.5 – Distribuição das variáveis nas planilhas de produtividade atualmente

Tipos de variáveis	Quantidade de variáveis na planilha	% em relação ao nº total
Publicação	39	36%
Impacto das publicações	19	18%
Formação de Recursos humanos em CT&I	21	19%
Gestão de projetos em CT&I	17	16%
Outras	12	11%

Fonte: CNPq. Elaboração: autora.

Durante a etapa de análise exploratória de dados, verificou-se que quase todas as variáveis referentes aos dados de produtividade em Ciência e Tecnologia dos pesquisadores poderiam ser agrupadas em quatro dimensões: 1) esforço de publicação; 2) impacto das publicações; 3) esforço de formação de recursos humanos; e 4) esforço de obtenção de capital científico (conceito abordado no capítulo 2), cujas variáveis são apresentadas nos Quadros 3.6 a 3.9.

No Quadro 3.6 temos a descrição das variáveis referentes ao esforço de publicação.

Quadro 3.6 - Dimensão Esforço de Publicação - Planilha de Produtividade (2013, 2014, 2017 e 2020)

Dimensão esforço de publicação	
Variável	Descrição
ARTIGOS_TOTAL	Nº de artigos publicados em periódicos
ART_JCR_MAIOR_IGUAL_6	Nº de artigos publicados em periódicos JCR >= 6
ART_JCR_5_E_MEIO_ATE_6	Nº de artigos publicados em periódicos JCR entre 5,5 e 6
ART_JCR_5_ATE_5_E_MEIO	Nº de artigos publicados em periódicos JCR entre 5 e 5,5
ART_JCR_4_E_MEIO_ATE_5	Nº de artigos publicados em periódicos JCR entre 4,5 e 5
ART_JCR_4_ATE_4_E_MEIO	Nº de artigos publicados em periódicos JCR entre 4 e 4,5
ART_JCR_3_E_MEIO_ATE_4	Nº de artigos publicados em periódicos JCR entre 3,5 e 4
ART_JCR_3_ATE_3_E_MEIO	Nº de artigos publicados em periódicos JCR entre 3 e 3,5
ART_JCR_2_E_MEIO_ATE_3	Nº de artigos publicados em periódicos JCR entre 2,5 e 3
ART_JCR_2_ATE_2_E_MEIO	Nº de artigos publicados em periódicos JCR entre 2 e 2,5
ART_JCR_1_E_MEIO_ATE_2	Nº de artigos publicados em periódicos JCR entre 1,5 e 2
ART_JCR_1_ATE_1_E_MEIO	Nº de artigos publicados em periódicos JCR entre 1 e 1,5
ART_JCR_MEIO_ATE_1	Nº de artigos publicados em periódicos JCR entre 0,5 e 1
ART_JCR_ATE_MEIO	Nº de artigos publicados em periódicos JCR entre 0 e 0,5
ART_JCR_NULO	Nº de artigos publicados em periódicos
ART_MENOR_IGUAL_10_AUTORES	Nº de artigos publicados em periódicos com 10 ou menos autores
ART_ENTRE_11_19_AUTORES	Nº de artigos publicados em periódicos com 11 a 19 autores
ART_MAIOR_IGUAL_20_AUTORES	Nº de artigos publicados em periódicos com 20 ou mais autores

ART_PRIMEIRO_AUTOR	Nº de artigos publicados em periódicos como primeiro autor
ART_ULTIMO_AUTOR	Nº de artigos completos publicados como último autor - somente se tiver mais de 1 autor
ART_JCR_1_3_PRIM_OU_ULT_AUT	Nº de artigos publicados em periódicos - JCR maior 1 e menor 3 informando entre eles quanto o solicitante é primeiro ou último autor
ART_JCR_1_3_SEG_OU_PENULT_AUT	Nº de artigos publicados em periódicos - JCR maior 1 e menor 3 informando entre eles quanto o solicitante é segundo ou penúltimo autor
ART_JCR_MAIOR_3_PRI_OU_ULT_AUT	Nº de artigos publicados em periódicos - JCR maior ou igual a 3 informando entre eles quanto o solicitante é primeiro ou último autor
ART_JCR_MAIOR_3_SEG_OU_PEN_AUT	Nº de artigos publicados em periódicos - JCR maior ou igual a 3 informando entre eles quanto o solicitante é segundo ou penúltimo autor
TRAB_PUBLICADOS_EVENTOS	Nº de trabalhos completos publicados em eventos
LIVROS	Nº de livros
CAPITULOS_LIVROS	Nº de capítulos de livros publicados
ORGANIZACAO_OBRA	Nº de obras organizadas
ART_QUALIS_A1	Nº de artigos publicados em periódicos Qualis A1
ART_QUALIS_A2	Nº de artigos publicados em periódicos Qualis A2
ART_QUALIS_B1	Nº de artigos publicados em periódicos Qualis B1
ART_QUALIS_B2	Nº de artigos publicados em periódicos Qualis B2
ART_QUALIS_B3	Nº de artigos publicados em periódicos Qualis B3
ART_QUALIS_B4	Nº de artigos publicados em periódicos Qualis B4
ART_QUALIS_B5	Nº de artigos publicados em periódicos Qualis B5
ART_QUALIS_C	Nº de artigos publicados em periódicos Qualis
ART_QUALIS_NULO	Nº de artigos publicados em periódicos Qualis não informado

Fonte: CNPq. Elaboração: Autora.

No Quadro 3.7 temos a descrição das variáveis referentes à Dimensão 2 -impacto das publicações. Ao todo existem 19 variáveis referentes a esta dimensão, conforme mostrado no Quadro 3.7.

Quadro 3.7 - Dimensão Impacto das Publicações - Planilha de Produtividade (2013, 2014, 2017 e 2020)

Dimensão 2: impacto das publicações	
Variável	Descrição
INDICE_H_ISI_INFORMADO	Índice H informado ISI
MEDIA_H_ISI_INF_NIVEL	Número médio de H em cada nível de bolsa - ISI Informado (Média calculada daqueles que tem H maior que 0)
INDICE_H_ISI_CALCULADO	Índice H Calculado ISI
MEDIA_H_ISI_CAL_NIVEL	Número médio de H em cada nível de bolsa - ISI Calculado (Média calculada daqueles que tem H maior que 0)
INDICE_H_SCOPUS_INFORMADO	Índice H informado Scopus
MEDIA_H_SCOPUS_INF_NIVEL	Número médio de H em cada nível de bolsa - SCOPUS Informado (Média calculada daqueles que tem H maior que 0)
INDICE_H_SCOPUS_CALCULADO	Índice H Calculado Scopus
MEDIA_H_SCOPUS_CAL_NIVEL	Número médio de H em cada nível de bolsa - SCOPUS Calculado (Média calculada daqueles que tem H maior que 0)
CITACOES_ISI_INFORMADO	Total de Citações informado ISI
MEDIA_CIT_ISI_INF_NIVEL	Número médio de citações em cada nível de bolsa - ISI Informado (Média calculada daqueles que tem Citações maior que 0)
CITACOES_ISI_CALCULADO	Total de Citações Calculado ISI
MEDIA_CIT_ISI_CAL_NIVEL	Número médio de citações em cada nível de bolsa - ISI Calculado (Média calculada daqueles que tem Citações maior que 0)
CITACOES_SCOPUS_INFORMADO	Total de Citações informado Scopus
MEDIA_CIT_SCOPUS_INF_NIVEL	Número médio de citações em cada nível de bolsa - SCOPUS Informado (Média calculada daqueles que tem Citações maior que 0)
MEDIA_CIT_SCOPUS_CAL_NIVEL	Número médio de citações em cada nível de bolsa - SCOPUS Calculado (Média calculada daqueles que tem Citações maior que 0)
CITACOES_SCOPUS_CALCULADO	Total de Citações Calculado Scopus
SOMATORIO_GERAL_JCR	Somatório de JCR artigos completos publicados
MEDIA_GERAL_JCR	JCR médio de artigos publicados em periódicos
MEDIA_AUTORES_JCR_MAIOR_1	Nº médio de autores em artigos publicados em periódicos com JCR maior que 1

Fonte: CNPq. Elaboração: Autora.

No Quadro 3.8 temos as variáveis referentes à dimensão 3 (Esforço de formação de recursos humanos). Ao todo existem 21 variáveis referentes a esta dimensão, conforme mostrado no Quadro 3.8.

Quadro 3.8 - Dimensão Esforço de Formação de Recursos Humanos - Planilha de Produtividade (2013, 2014, 2017 e 2020)

Dimensão 3: esforço de formação de recursos humanos	
Variável	Descrição
ORIENT_IC_CONCLUIDOS	Nº de orientações em iniciação científica concluídos
ORIENT_GM_CONCLUIDOS	Nº de orientações de mestrado concluídas
ORIENT_GD_CONCLUIDOS	Nº de orientações de doutorados concluídas
SUPERV_PD_CONCLUIDOS	Nº de supervisão de pós-doutorados concluídas
ORIENT_IC_ANDAMENTO	Nº de orientações em iniciação científica em andamento
ORIENT_GM_ANDAMENTO	Nº de orientações de mestrado em andamento
ORIENT_GD_ANDAMENTO	Nº de orientações de doutorados em andamento
SUPERV_PD_ANDAMENTO	Nº de supervisão de pós-doutorados em andamento
ORIENT_IC_CONCLUIDOS_TOT	Nº de orientações em iniciação científica concluídos - total do pesquisador
ORIENT_GM_CONCLUIDOS_TOT	Nº de orientações de mestrado concluídas - total do pesquisador
ORIENT_GD_CONCLUIDOS_TOT	Nº de orientações de doutorados concluídas - total do pesquisador
SUPERV_PD_CONCLUIDOS_TOT	Nº de supervisão de pós-doutorados concluídas - total do pesquisador
CO_ORIENT_GM_CONCLUIDOS	Nº de coorientações de mestrado concluídas
CO_ORIENT_GD_CONCLUIDOS	Nº de coorientações de doutorados concluídas
CO_ORIENT_GM_ANDAMENTO	Nº de coorientações de mestrado em andamento
CO_ORIENT_GD_ANDAMENTO	Nº de coorientações de doutorados em andamento
CO_ORIENT_GM_CONCLUIDOS_TOT	Nº de coorientações de mestrado concluídas - total do pesquisador
CO_ORIENT_GD_CONCLUIDOS_TOT	Nº de coorientações de doutorados concluídas - total do pesquisador
ORIENT_TCC_CONCLUIDAS	Nº de orientações de TCC concluídas

Fonte: CNPq. Elaboração: Autora.

No Quadro 3.9 temos as variáveis referentes à Dimensão 4 (esforço de obtenção de capital científico). Ao todo 17 variáveis referentes a esta dimensão, conforme mostrado no quadro 3.9.

Quadro 3.9 - Dimensão Esforço de obtenção de capital científico - Planilha de Produtividade (2013, 2014, 2017 e 2020)

Dimensão 4: Esforço de obtenção de capital científico	
Variável	Descrição
MEMBRO_CORPO_EDITORIAL	Nº de periódicos onde é membro de corpo editorial (Atual)
EVENTOS_ORGANIZADOR	Nº de eventos nos quais é organizador
EVENTOS_PALESTRANTE_CONVIDADO	Nº de eventos onde é participante convidado (Conferencista ou Simposista)
PARTICIP_COMITE_ASSESS	Nº de comitês de assessoramento onde participa
PREMIOS_E_TITULOS	Número de prêmios e títulos honoríficos recebidos
PROJETOS_CERTIFICADOS	Número de projetos certificados no Currículo Lattes
PARTICIP_BANCA_DOUT_CONC	Nº de participação em bancas de doutorado concluídas
PARTICIP_BANCA_MESTR_CONC	Nº de participação em bancas de mestrado concluídas
COORD_PROJETO_PESQUISA	Nº de coordenações de projetos de pesquisa
MEMBRO_EQUIPE_PROJ_PESQ	Nº de membros de equipe de projeto de pesquisa
COORD_PROJETO_DESENV	Nº de coordenações de projeto de desenvolvimento
MEMBRO_EQUIPE_PROJ_DESENV	Nº de membros de equipe de projeto de desenvolvimento
COORD_PROJETO_EXTENSAO	Nº de coordenações de projeto de extensão
MEMBRO_EQUIPE_PROJ_EXTENSAO	Nº de membros de equipe de projeto de extensão
PARTICIP_BANCA_DOUT_CONC	Nº de participação em bancas de doutorado concluídas

Fonte: CNPq. Elaboração: Autora.

3.3 Utilizando Análise de Componentes Principais para Redução de Dimensionalidade

Nos Quadros 3.6 a 3.9 foram apresentados todos os indicadores presentes na planilha de produtividade, os quais foram agrupados em quatro dimensões, somando 96 variáveis. Sendo assim, optou-se pela redução do número de variáveis utilizando a Análise de Componentes Principais (ACP) ou PCA (*Principal Component Analysis*) como recurso matemático para diminuir o número de variáveis a serem utilizadas. Esta técnica permite que se reoriente os dados “de modo que as primeiras poucas dimensões expliquem o maior número possível de informações disponíveis” (LATTIN *et al*, 2011, p. 67).

A ideia central da PCA é reduzir a dimensionalidade de um conjunto de dados, que possivelmente possuem um grande número de variáveis correlacionadas, por meio da transformação dos dados em um novo conjunto de variáveis, chamadas de componentes principais, que nada mais são do que combinações lineares das variáveis originais e estas combinações lineares são não correlacionadas entre si (MINGOTI, 2005). Assim, a análise fatorial por componentes principais permite, que, a partir de um conjunto de variáveis originais correlacionadas entre si, “seja determinado novo conjunto de variáveis (fatores) resultantes da combinação linear do primeiro conjunto” (FÁVERO; BELFIORE, 2017, p. 409).

Se temos p -variáveis originais é possível a obtenção de p componentes principais. No entanto, o intuito, em geral é a redução do número de variáveis a serem avaliadas, ou seja, a informação contida nas p -variáveis originais é substituída pela informação contida em k componentes principais, sendo $k < p$ (MINGOTI, 2005). Os componentes obtidos com esta técnica, são ordenados de forma que os primeiros guardam a maior parte da variância presente nas variáveis originais. Desta forma, é possível reduzir a dimensão do conjunto de componentes principais, descartando as componentes com menor variabilidade.

Segundo Mingoti (2005), a obtenção das componentes principais envolve a decomposição da matriz de covariâncias do vetor de interesse e caso seja feito algum tipo de transformação deste vetor de interesse, as componentes deverão ser determinadas com base no vetor transformado. Uma transformação bastante comum é a padronização das variáveis, que resulta em uma média igual a 0 e um desvio padrão igual a 1.

Segundo Nurfalih, Adiwijaya e Ardiyanti (2016), os componentes principais são extraídos por meio de autovalores e autovetores que podem ser obtidos com os cálculos a seguir.

Sejam os dados da matriz X (m, n), cada m -amostra na matriz X é representada como um vetor de tamanho n . A matriz X é a entrada do PCA. Primeiro, determina-se a média dos dados:

$$\bar{X} = \sum_{i=1}^n \frac{X_i}{n} \quad (3.1)$$

Sendo X_i as variáveis de entrada e n o tamanho da entrada. Após encontrada a média é possível calcular a matriz de covariância C_x :

$$C_x = \sum_{i=1}^n \frac{(X_i - \bar{X}) - (X_i - \bar{X})^T}{n-1} \quad (3.2)$$

A partir da matriz de covariância C_x é possível calcular os autovalores e autovetores com (3.3):

$$CU_n = \lambda_n U_n \quad (3.3)$$

Onde U é o autovetor e λ o autovalor. Com isso é possível realizar a transformação dos dados utilizando a fórmula (4):

$$Y = U^T * (X - \bar{X}) \quad (3.4)$$

O uso de PCA⁸ permitiu a construção de quatro indicadores, cada um referente a uma dimensão e tais indicadores foram inseridos na base de dados criada. O recorte utilizado na base de dados é referente aos anos de 2013, 2014, 2017 e 2020, como já mencionado.

Além do grande número de variáveis, também existe o fato de que cada comitê utiliza os indicadores de produtividade de forma diferente. O que é importante num comitê não é em outro e isso dificulta a análise. Reunir todas as variáveis referentes às dimensões em seus respectivos índices, torna a análise mais simples e facilita a construção do modelo.

Conforme já mencionado, para cada dimensão foi elaborado um indicador:

⁸ Os valores das componentes principais foram obtidos utilizando o software R. Há muitas formas de fazer a PCA em R, em diversos pacotes distintos. Neste trabalho foi utilizado a função *prcomp* do pacote *stats* no R (versão 4.1.2).

- 1) IND_P – Indicador do esforço de publicação do pesquisador, reúne os indicadores de publicação em artigos, livros, capítulos de livros, organização de obra e trabalhos em eventos.
- 2) IND_IMP – Indicador do impacto das publicações de cada pesquisador. Este indicador reúne as variáveis referentes ao impacto das publicações disponibilizados na planilha de produtividade.
- 3) IND_RH – Indicador do esforço e da capacidade do pesquisador em gerar recursos humanos por meio da orientação e/ou coorientação de trabalhos de conclusão de curso, mestrado, doutorado e supervisão de pós-doutorado.
- 4) IND_CC – Indicador do esforço de obtenção de capital científico, que aqui se configura como o esforço do pesquisador na gestão de projetos de pesquisa, desenvolvimento e extensão seja como coordenador dos projetos ou como membro da equipe. Este indicador também reúne variáveis referentes ao reconhecimento do pesquisador pelos seus pares, como ser membro de comitê, participação em eventos como convidado, prêmios e títulos etc. Conforme mostrado no quadro 3.9.

A metodologia de construção dos 04 indicadores, referentes às quatro dimensões mencionadas seguiu os seguintes passos⁹:

1. Foram carregados os dados do banco criado;
2. Foram selecionadas as variáveis de interesse de cada dimensão e foram utilizadas apenas as variáveis presentes na planilha de produtividade de 2013, para não afetar o resultado da regressão;
3. As variáveis foram separadas por ano;
4. Os dados referentes a cada variável foram padronizados ano a ano;
5. Com a técnica de PCA (*Principal Component Analysis*) foram obtidas as componentes principais para cada ano separadamente;
6. Os vetores das primeiras componentes obtidas foram reunidas num único *dataframe* de X linhas por 1 coluna;
7. Os indicadores criados foram escalonados numa escala de 0 a 100, para facilitar a análise;

⁹ Todas estas etapas foram feitas no software R, versão 4.1.2 e a programação em R está apresentada no Apêndice B desta dissertação. Os indicadores foram criados utilizando-se a primeira componente obtida (PC1).

8. O *dataframe* criado foi introduzido no bando de dados como uma nova variável de mesmo nome do indicador (IND_P, IND_IMP, IND_RH e IND_CC).

No banco de dados, as variáveis criadas foram mantidas com seus nomes originais e as mesmas variáveis escalonadas foram acrescentadas com o final N (IND_PN, IND_IMP_N, IND_RHN e IND_CCN).

No Quadro 3.10 estão mostrados os percentuais de explicação da variabilidade dos dados para cada indicador criado.

Quadro 3.10 – Percentual da variabilidade explicada pela componente 1 (PC1)

Indicador	Explicação da variabilidade dos dados (PC1) em %			
	2013	2014	2017	2020
IND_P	24,13	20,72	23,17	29,13
IND_IMP	64,17	66,83	70,19	65,78
IND_RH	31,43	30,14	26,20	26,22
IND_CC	33,92	29,93	32,29	30,42

Elaboração: Autora.

3.4 Escolha dos Grupos de Tratamento e Controle

O presente estudo analisa o efeito na carreira de pesquisadores antes e depois da obtenção de uma bolsa de produtividade PQ. Para isso foi utilizado o modelo de dados em painel com estimador de diferenças em diferenças, que possibilita uma avaliação dos efeitos de uma bolsa PQ na carreira acadêmica dos pesquisadores brasileiros. O estudo abrange 12 comitês, conforme já descrito em tópicos anteriores e utiliza os dados de 2013, 2014, 2017 e 2020.

O número de pesquisadores que submetem propostas, anualmente, varia de 8 a 11 mil pesquisadores ao ano, vide Quadro 3.11. Na coluna 3 do Quadro 3.11, temos a quantidade de propostas submetidas nos anos de 2013, 2014, 2017 e 2020, nos 12 comitês estudados nesta pesquisa. Na coluna 4, temos o percentual de propostas submetidas nestes comitês em relação ao número total de propostas. O número de propostas submetidas está compatível com o que foi apresentado no Quadro 3.4, no qual é apresentado a distribuição do número de bolsas por

comitê e onde foi destacado que os 12 comitês estudados possuem cerca de 44% das bolsas PQ ativas.

Quadro 3.11 – Número Total de Propostas submetidas nas Chamadas de Bolsas PQ em 2013, 2014, 2017 e 2020

ANO	Total de propostas submetidas (todos os comitês) (1)	Total de propostas submetidas (apenas os comitês estudados) (2)	(%) Total (2) / (1)
2013	9766	4033	41,3
2014	8974	3914	43,6
2017	10199	4006	39,3
2020	11596	3936	33,9

Elaboração: Autora. Fonte: CNPq.

O ano de 2013 está sendo utilizado como referência, pois, conforme já explicitado neste documento, as planilhas de produtividade começaram a ser produzidas de forma institucional neste ano. O ano de 2013 é, portanto, o ano anterior à política, ou seja, é o ano no qual, nenhum dos indivíduos selecionados foi contemplado com bolsa PQ. Em 2013 foram selecionados todos os indivíduos que obtiveram parecer de recomendação dos comitês analisados, mas que não foram aprovados nas respectivas diretorias (falta de cotas), ou seja, todos esses indivíduos começam em 2013 sem bolsa PQ. Desta forma, foram selecionados 645 indivíduos nesta condição, em 2013.

Uma bolsa PQ-2 tem a duração de 36 meses, sabendo que os pesquisadores que não possuem bolsa PQ continuam usualmente tentando, ano após ano, nesta pesquisa é acompanhado ao longo do tempo, quantos destes indivíduos continuaram tentando obter uma bolsa em 2014, 2017 e 2020. Assim, é possível acompanhar quais destes indivíduos tornaram-se bolsistas PQ-2 em 2014 e/ou em 2017 e 2020, comparando suas trajetórias acadêmicas com os indivíduos que não obtiveram bolsa no mesmo período. A ideia é a de que os pesquisadores que obtiveram bolsa PQ-2 em 2014, provavelmente submeteram propostas (concorreram novamente a uma bolsa PQ) em 2017 e 2020, e indivíduos que não conseguiram bolsa em 2014, vão continuar tentando.

Durante a análise exploratória dos dados, verificou-se que dos 645 indivíduos selecionados em 2013, 371 submeteram propostas em 2014, 184 em 2017 e 151 ainda permaneciam em 2020. Em 2014 foi verificado que 98 dos 371 indivíduos foram contemplados com bolsa PQ-2. Esse grupo de 98 pesquisadores contemplados com bolsa PQ-2 em 2014, os quais também haviam enviado solicitação em 2013, mas não foram contemplados, foram caracterizados como o grupo de tratamento (T) e os demais como grupo de controle (C). Na

base de dados, foi criada uma nova variável chamada de “GRUPO”, onde os pesquisadores do grupo de tratamento recebem o valor “T” e os do grupo de controle valor “C”.

No Quadro 2.12, temos o resultado final do painel de dados obtidos. Ao todo, o painel possui 1351 registros de pesquisadores, alguns aparecem 04 vezes ao longo do painel, outros 03 vezes, outros 02 vezes e outros apenas 01. O painel final, trata-se, portanto, de um painel desbalanceado, pois os períodos de tempo analisados não possuem o mesmo número de observações. A seguir é apresentado a caracterização do painel obtido quanto ao número de pesquisadores.

Quadro 3.12 – Resultado do Painel obtido em número de indivíduos

Grupos	2013 (antes do tratamento)	2014 (início do tratamento)	2017 (03 anos após início do tratamento)	2020 (06 anos após início do tratamento)
Tratamento (T)	98	98	77	56
Controle (C)	547	273	107	95
Total	645	371	184	151

Fonte: CNPq, 2020. Elaboração: Autora.

Em 2014, conforme já salientado, 98 pesquisadores, em um total de 371, foram contemplados com bolsa PQ, tais indivíduos foram classificados como grupo de tratamento e são acompanhados ao longo do anos de 2017 e 2020. Podemos verificar no Quadro 3.12, que o número de pesquisadores no grupo de controle e grupo de tratamento diminuem ao longo dos anos. Sendo a redução muito maior no grupo de controle (547 inicialmente e 95 ao final). O que ocorre, é que conforme já discutido neste documento, estes pesquisadores ao não serem contemplados com bolsa PQ, continuam tentando nos anos subsequentes, tendo em vista a importância que essa bolsa possui no universo da pesquisa brasileira. Muitos destes tentaram a obtenção de bolsa PQ em 2015 e 2016 e muitos obtiveram esse objetivo e assim também em 2018 e 2019. Além disso, durante a análise exploratória de dados, verificou-se que há uma grande mobilidade entre os pesquisadores. Dos 645 pesquisadores presentes na amostra inicial, 98 obtiveram bolsa em 2014. Destes, 12 progrediram de PQ-2 para PQ-1D em 2017 e assim, saem da amostra em 2020, pois a bolsa PQ-1D tem duração de 48 meses. Mesmo assim, dos 98 pesquisadores que foram contemplados com bolsa PQ-2 em 2014, apenas 77 permanecem na amostra em 2017. Ou seja, 21 pesquisadores saíram da amostra em 2017. Duas possibilidades

podem ter ocorrido: a) uma parte destes pesquisadores podem ser mulheres que pediram prorrogação parturiente (12 meses) por parto ocorrido no período da bolsa e assim, só vão submeter pedido novamente em 2018; ou b) se trata de uma questão de esquecimento de submissão, que é relativamente comum. Não há a possibilidade destes pesquisadores terem enviado proposta em 2015 e 2016 e sido contemplados com bolsa PQ-1D, porque isso não é permitido, nas normas de bolsas PQ. O pesquisador não pode solicitar uma bolsa PQ, se tem uma bolsa vigente, com data de encerramento superior ao dia 28 de fevereiro do ano subsequente à submissão da proposta.

Ainda com respeito ao grupo de tratamento, de 2017 para 2020, esse grupo foi reduzido de 77 para 56 pesquisadores. Obviamente, os 12 pesquisadores que foram contemplados com bolsa PQ-1D não iriam aparecer de qualquer forma em 2020, mas ainda assim deveria haver 65 e não 56. A diferença de nove pesquisadores pode estar ligada às duas possibilidades já citadas anteriormente (prorrogação parturiente ou esquecimento).

Quanto ao grupo de controle, durante a fase de análise exploratória de dados, verificou-se que em 2017, todos os 107 indivíduos do grupo de controle foram contemplados com bolsa PQ, destes 95 enviaram solicitação de bolsa em 2020, ou seja, 12 pesquisadores deixaram de submeter proposta (pode ter sido por prorrogação parturiente ou esquecimento). Destes 95, que haviam sido contemplados com bolsa PQ-2 em 2017 e submeteram novamente em 2020, 47 perderam a bolsa em 2020 e 48 continuaram com bolsa PQ-2. No entanto, isso não significa que 49% dos pesquisadores, pertencentes ao grupo de controle perderam a bolsa PQ de um julgamento para outro, já que foi verificado que não se trata dos mesmos pesquisadores. Pois, pesquisadores que foram contemplados com bolsa em 2020, não estavam no mesmo grupo de pesquisadores, pertencentes ao grupo de controle, que foram contemplados com bolsa em 2017. E há pesquisadores, que foram contemplados em 2017 e perderam bolsa em 2020. Enquanto no grupo de controle, em 2020, 07 pesquisadores perderam suas bolsas e 06 progrediram para o nível 1D.

Do que foi relatado até aqui é preciso salientar, que dentro o grupo de controle há bastante mobilidade, pois pesquisadores do grupo de controle receberam a política em 2017 ou 2020.

3.4.1 Caracterização dos Grupos de Tratamento e Controle

No Quadro 3.13, temos a distribuição do painel por gênero e no Quadro 2.14, a distribuição por região.

Quadro 3.13 – Distribuição do Painel por gênero

Ano	Gênero					
	Feminino	Masculino	Total	Feminino (%)	Masculino (%)	Total (%)
2013	234	411	645	36,3	63,7	100
2014	126	245	371	34,0	66,0	100
2017	67	117	184	36,4	63,6	100
2020	49	102	151	32,4	67,6	100

Fonte: CNPq, 2020. Elaboração: Autora.

No Quadro 3.13, pode-se notar que nos anos em análise, entre 32 a 36% dos indivíduos da amostra são mulheres, o que está de acordo com o que foi analisado no Capítulo anterior, no que diz respeito ao número total de pesquisadores com bolsa PQ no CNPq.

No Quadro 3.14, temos a distribuição do painel por região.

Quadro 3.14 – Distribuição do Painel por região

Ano	Região						
	Centro-oeste	Nordeste	Norte	Sudeste	Sul	Total	(%) Sul e Sudeste
2013	58	112	20	348	107	645	70,5
2014	30	62	9	202	68	371	72,8
2017	16	21	3	109	35	184	78,2
2020	13	18	2	93	25	151	78,1
Total	117	213	34	752	235	1351	73%

Fonte: CNPq, 2020. Elaboração: Autora.

No Quadros 3.14, pode-se perceber que no período analisado 72 a 78% são indivíduos das regiões Sul e Sudeste. Norte, Nordeste e Centro-oeste contemplam de 22 a 28% dos indivíduos da amostra. Sendo a região Norte, a região com menor número de pesquisadores na amostra. O que está de acordo com o que foi relatado no Capítulo anterior, onde a distribuição de bolsas PQ por região foi apresentada no Gráfico 1.8 e onde pudemos ver que a região Sudeste

concentra 58% de todas as bolsas PQ em atividade, seguida pela Sul com 20%. Ou seja, as regiões Sul e Sudeste concentram 78% das bolsas PQ ativas do CNPq.

No Quadro 3.15, temos a distribuição do painel por comitê. Dentre os 12 comitês analisados, apenas 4, não configuram entre os 10 maiores comitês do CNPq: ZO – Zoologia, VT - Medicina Veterinária, EE - Engenharias Elétrica e Biomédica e MM - Engenharias de Minas e de Metalúrgica e Materiais. Pode-se perceber a predominância dos comitês de Física e Astronomia (FA) e Química (QU) entre os comitês analisados. O Comitê de Física e Astronomia é o maior comitê do CNPq, seguido pelos comitês de Biofísica, Bioquímica, Farmacologia, Fisiologia e Neurociências (BF), Agronomia (AG) e Química (QU). No entanto, isso não aparece refletido na amostra.

Quadro 3.15 – Distribuição do Painel por Comitê

ANO	Comitês Analisados											
	AG	BF	CA	CS	EE	FA	LL	MD	MM	QU	VT	ZO
2013	4	21	73	82	4	152	66	9	42	137	18	37
2014	6	13	36	49	4	89	42	3	28	74	11	16
2017	2	9	18	30	4	47	17	0	21	26	6	4
2020	0	0	9	19	4	39	11	0	15	42	3	3

Fonte: CNPq, 2020. Elaboração: Autora.

4. ESTRATÉGIA EMPÍRICA

4.1 Modelo de Regressão de Dados em Painel

Neste trabalho, a metodologia implementada utiliza a técnica de dados em painel com o estimador de diferenças em diferenças para analisar o efeito de uma bolsa de produtividade PQ, na carreira de pesquisadores brasileiros ao longo dos anos de 2013, 2014, 2017 e 2020. Isso é possível, quando se dispõe de informações sobre unidades, indivíduos, cidades, empresa ou países, não apenas para um único instante de tempo (uma única *cross-section*), mas em vários períodos de tempo (várias *cross-sections*, como semanas, meses, trimestres ou anos) (MALBOUISSON; TIRYAKI, FÁVERO; BELFIORE, 2017). Particularmente, são apresentadas as estimações por Mínimos Quadrados Ordinários (MQO) para dados em painel (*pooled OLS*), modelo de efeitos fixos e modelo de efeitos aleatórios.

Gujarati e Porter (2011) enumeram algumas vantagens e desvantagens da utilização de dados em painel. Uma das principais vantagens da utilização de modelos de dados em painel consiste em possibilitar o estudo das diferenças existentes em determinado fenômeno entre indivíduos em cada *cross-section*, além de permitir a análise da evolução temporal deste mesmo fenômeno para cada indivíduo. Os dados em painel são mais adequados para estudar a dinâmica da mudança (GUJARATI; PORTER, 2011). Além disso, os modelos longitudinais de regressão providenciam maior quantidade de informação, maior variabilidade dos dados, uma menor multicolinearidade entre as variáveis, maior número de graus de liberdade e maior eficiência quando da estimação de seus parâmetros (GUJARATI; PORTER, 2011, FÁVERO; BELFIORE, 2017).

Como limitações do uso de dados em painel, principalmente quando se trata de painéis com informações de indivíduos, são problemas com o levantamento de dados obtidos em entrevistas, como cobertura incompleta da população em estudo, ausência de cooperação, período de referências, erros de preenchimento, dentre outros. Além de problemas com dimensão temporal restrita (poucos períodos de tempo); erros de medidas de variáveis, devido a perguntas pouco claras, ou problemas de compreensão, problemas de seletividade, relacionados à auto seleção, ausência de respostas e atrito. O atrito ocorre quando há problema de perda das unidades amostrais ao longo do tempo. Pode haver ainda problemas com dependência de *cross-section* devido à correlação entre unidades individuais (GUJARATI; PORTER, 2011, MALBOUISSON; TIRYAKI, FÁVERO; BELFIORE, 2017).

Os dados em um painel são ditos balanceados, quando cada unidade de corte transversal (países, empresas, indivíduos etc.) tem o mesmo número de observações. Se cada unidade em análise tiver um número diferente de observações, é dito que o painel é desbalanceado. Os painéis ainda podem curtos ou longos. Em um painel curto, o número de unidades do corte transversal, N , é maior que o número de períodos de tempo T , em análise. Em um painel longo, o número de períodos de tempo T é maior que N (GUJARATI: PORTER, 2011).

4.1.1 Notação e tipos de regressão de dados em painel

A notação clássica do modelo de regressão linear, assumindo dados apenas transversais é dado pela equação 4.1:

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad (4.1)$$

na qual Y representa a variável dependente, α indica o intercepto, β representa a variação observada em Y quando a variável independente, X , aumenta em uma unidade, e ε indica o termo de erro. O subscrito i indica que as observações são indexadas por caso.

A regressão de painel é uma extensão do modelo anterior, em que:

$$Y_{it} = \alpha_i + \beta X_{it} + \varepsilon_{it} \quad (4.2)$$

Os subscritos i e t informam que as observações são indexadas, respectivamente, por caso e tempo. Para prosseguir com as regressões, vamos apresentar três abordagens tradicionais para estimar modelos de painel.

4.1.2 Modelo MQO para dados empilhados (*pooled data*)

Essa abordagem pressupõe que todas as unidades podem ser contempladas pela mesma reta de regressão ao reunir todas as observações em uma mesma regressão linear utilizando Mínimos Quadrados Ordinários (MQO). Mesmo que as unidades i apresentem diferenças umas das outras, considera-se que o conjunto de variáveis independentes já carrega toda a informação importante para o modelo. Isso pressupõe que os principais fatores que diferenciam as observações já estão explicitados no modelo, de modo que não é necessário controlar por outros fatores não-observados.

Como alerta Malbouisson e Tityaki (2017), na presença de efeitos individuais ou temporais essa abordagem viola dois pressupostos MQO: 1) presume que os erros são homocedásticos, ou seja, possuem a mesma variância; e 2) presume que os erros não são correlacionados entre si. No entanto, na presença de efeitos individuais ou temporais é bastante comum a existência de heterocedasticidade, ou seja, variâncias diferentes entre as observações individuais ou variâncias diferentes entre observações temporais. Além disso, é possível que erros sejam correlacionados para cada unidade individual.

Assim, nessa abordagem, há apenas um intercepto (α) para toda a população e todo o mais que não é observável é alocado como fator de erro (ε). Isso quer dizer que se houver fatores não-observados em ε que não sejam aleatórios, essa abordagem já não seria apropriada. Se tivermos elementos não-observados correlacionados com X , temos o que se chama de viés de variável omitida e desta forma, a estimativa do parâmetro β não é consistente (GUJARATI; PORTER, 2011).

Tais fatores não-observados se dividem em elementos que permanecem inalterados para cada unidade e elementos que variam segundo a unidade e o tempo. Desta forma, podemos decompor o erro da Equação (4.2) em duas partes: uma fixa para cada unidade (μ_i) e outra variável (v_{it}):

$$\varepsilon_{it} = \mu_i + v_{it} \tag{4.3}$$

O termo (μ_i) representa efeito individual que não é observado e que não se modifica ao longo do tempo, enquanto v_{it} captura a parte do termo de erro que é independente, também chamado de efeito variável, erro de variação temporal ou idiossincrático (WOOLDRIDGE, 2016).

Vale salientar que o modelo de regressão MQO para dados empilhados em painel não dispõe de estratégias para lidar com μ_i , pois não desagrega ε_{it} . Assim, a seguir são apresentadas outras duas abordagens que decompõem o fator de erro em fixo e variável.

4.1.3 Modelo de Efeitos Fixos

No modelo de efeitos fixos, a possibilidade de que os interceptos variam entre as unidades individuais de *cross-section* ou entre as observações temporais é considerada e assume-se que os coeficientes estimados para as variáveis independentes são similares para as unidades individuais ou temporais (MALBOUISSON; TIRYAKI, 2017). O modelo de efeitos fixos leva em consideração a heterogeneidade entre as unidades observacionais, permitindo que cada uma tenha seu próprio intercepto. A denominação “efeitos fixos” decorre do fato de que este modelo controla os efeitos das variáveis não observadas que variam entre as unidades de corte transversal, mas são constantes ao longo do tempo. Dessa forma, supõe-se que o intercepto varia de um indivíduo para outro, mas é constante no tempo, e os parâmetros a serem estimados são constantes para os indivíduos.

O modelo de Efeito Fixo pode ser representado da seguinte forma:

$$Y_{it} = \alpha_i + \beta X_{it} + \varepsilon_{it} \quad (4.4)$$

Onde a principal característica deste modelo é tratar os α_i 's como variáveis aleatórias não observadas e correlacionadas com algum X_{it} .

Conforme apresentado em Wooldridge (2016), é possível estimar o modelo de efeitos fixos chamado de intragrupo ou *within effects*, modelo que é utilizado neste trabalho. Basicamente, o que este modelo faz é se utilizar de um procedimento para eliminar o efeito fixo, estimando a regressão com MQO utilizando variáveis ajustadas pela média de unidade *cross-section*. Segundo Fávero e Belfiore (2017), uma estimação *within* faz com que os dados sejam diferenciados em torno da média quando da modelagem e, como consequência, o parâmetro de determinada variável que apresentar dados que não sejam alterados ao longo do tempo não poderá ser estimado.

Para a regressão apresentada pela Equação (4.4), o modelo de efeitos fixos *within* consiste em estimar:

$$(Y_{it} - \bar{Y}_i) = \beta_1(X_{1it} - \bar{X}_{1i}) + \beta_2(X_{2it} - \bar{X}_{2i}) + \dots + \beta_k(X_{kit} - \bar{X}_{ki}) + (\varepsilon_{it} - \bar{\varepsilon}_i) \quad (4.5)$$

A vantagem desse modelo, segundo Malbouisson e Tiryaki (2017) é preservar o número de graus de liberdade e, se as variáveis em X forem independentes do termo de erro (exógenas), os estimadores não serão enviesados. No entanto, conforme já explicado, é que este procedimento de ajuste das variáveis pela média elimina variáveis explicativas que não variam na dimensão temporal.

4.1.4 Modelos de Efeitos Aleatórios

Na estimativa do modelo de efeitos aleatórios, utiliza-se o método dos mínimos quadrados generalizados (MQG), que é apropriado quando temos autocorrelação dos resíduos e erros heterocedásticos. Diferentemente do modelo de efeitos fixos, o modelo de efeitos aleatórios considera que os efeitos em *cross-section* e temporais são aleatórios.

O modelo de efeitos aleatórios pode ser expresso da seguinte forma:

$$Y_{it} = \alpha_i + \beta_0 + \beta_1 X_{it1} + \beta_2 X_{it2} + \dots + \beta_k X_{itk} + \varepsilon_{it}, \quad t = 1, 2, \dots, T \quad (4.6)$$

O modelo de efeitos aleatórios assume que α_i em (4.6) é puramente aleatório, implicando a ausência de correlação entre α_i e os regressores. Assim, a equação (4.6) torna-se um modelo de efeitos aleatórios quando se presume que o efeito não observado α_i é não correlacionado com cada variável explicativa em todos os períodos de tempo, ou seja, $Cov(X_{itj}, \alpha_i) = 0$, $t = 1, 2, \dots, T$; $j = 1, 2, \dots, k$. Caso esta hipótese seja satisfeita, podemos considerar o erro combinado, isto é, $v_{it} = \alpha_i + \varepsilon_{it}$. Assim, reescrevendo (4.6), temos:

$$Y_{it} = \beta_0 + \beta_1 X_{it1} + \beta_2 X_{it2} + \dots + \beta_k X_{itk} + v_{it}, \quad t = 1, 2, \dots, T \quad (4.7)$$

e pressupõe-se que α_i é IID (independente e identicamente distribuído), com variância σ_α^2 e que ε_{it} é IID com variância σ_ε^2 . Como α_i é o erro composto em cada período de tempo, os v_{it} são serialmente correlacionados ao longo do tempo, logo, sob as hipótese de efeitos aleatórios, $\rho_v = Corr(v_{it}, v_{is}) = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2}$, para todo $t \neq s$, em que $\sigma_\alpha^2 = Var(\alpha_i)$ e $\sigma_\varepsilon^2 = Var(\varepsilon_{it})$.

Conforme Wooldridge (2016), a presença de correlação serial no termo de erro composto v_{it} pode ser considerável, visto que os erros-padrão do MQO agrupado ignoram essa

correlação, eles serão viesados, como também serão enviesadas as estatísticas de testes habituais. A fim de resolver estes problemas, Wooldridge (2016) recomenda a utilização de estimadores de mínimos quadrados generalizados (MQG), que considera a correlação entre os erros de cada observação.

4.2 Modelo de Diferenças em Diferenças

O presente estudo analisa o efeito de uma bolsa de produtividade PQ, antes e depois de sua obtenção, na carreira de pesquisadores brasileiros. Assim, os dados de pesquisadores que submeteram propostas de bolsa PQ nos anos de 2013, 2014, 2017 e 2020 foram agrupados em um formato de painel. A metodologia aplicada neste estudo para determinar o efeito dessas bolsas de produtividade na carreira de pesquisadores brasileiros é o método de Diferenças em Diferenças (DID). A principal hipótese do modelo é que a trajetória ao longo do tempo da variável de efeito selecionada do grupo de controle consiga representar o contrafactual do que aconteceria com o grupo de tratamento, caso essa bolsa não existisse.

Conforme descrito na seção 4.4, é considerado como grupo de tratamento, os pesquisadores que obtiveram bolsa PQ-2 em 2014, nos 12 comitês em análise; o grupo de controle corresponde aos pesquisadores que não obtiveram bolsa neste mesmo ano.

Esse método consiste em estudar as diferenças entre esses dois grupos, sendo necessários dados anteriores e posteriores à política – obtenção de bolsa PQ – para os dois grupos. Neste trabalho foi utilizado o modelo de DID para verificar o efeito da bolsa de produtividade PQ em dois períodos distintos: 1) entre 2013 e 2017 (primeiro período do pesquisador que recebeu bolsa PQ); e b) entre 2013 e 2020 (dois períodos completos de bolsa PQ-2).

Assim, a amostra é dividida em quatro categorias: o grupo de controle antes da mudança, o grupo de controle depois da mudança, o grupo de tratamento antes da mudança e o grupo de tratamento depois da mudança. O Quadro 4.1 traz de maneira mais clara como é realizada esta manipulação dos dados.

Quadro 4.1 – Método de Diferenças em Diferenças utilizando Grupo de Tratamento e Grupo de Controle

Grupo	Antes	Depois	Diferenças
Controle	C_A	C_D	$C_D - C_A$
Tratamento	T_A	T_D	$T_D - T_A$
Diferenças	$T_A - C_D$	$T_A - T_D$	$(T_D - T_A) - (C_D - C_A)$

Elaboração: Própria.

O Quadro 4.1 traz de forma bastante simples, como o grupo de tratamento e o de controle mudaram ao longo do tempo. $C_D - C_A$ representa a mudança na carreira do pesquisador ocorrida entre o ano inicial de obtenção de bolsa PQ (2014) e o último ano de análise (2020), ou seja, esse índice capta as mudanças que se deveram a outros fatores (que não a obtenção da bolsa) que também devem ter influenciado o grupo de tratamento. De modo análogo, $T_D - T_A$ representa a mudança ocorrida ao longo do tempo no grupo de tratamento, ou seja, capta as mudanças na carreira desse grupo provocadas pela obtenção da bolsa PQ, bem como as que se deveram por outros fatores, os quais influenciaram também o grupo de controle.

Quando se subtrai “ $C_D - C_A$ de “ $T_D - T_A$ ”, encontra-se a diferença entre os grupos de tratamento e o grupo de controle observada depois do impacto da obtenção de uma bolsa PQ, ou de outra forma, encontra-se a diferença da diferença verificada entre os grupos, entre os dois períodos, daí porque recebe o do nome diferenças em diferenças. Portanto, esse valor capta as mudanças na carreira ocorridas para os indivíduos do grupo de tratamento provocadas por eventos que não atingiram o grupo de controle. O valor obtido pelo método de diferenças em diferenças pode ser representado por meio da expressão matemática (4.8):

$$\alpha = (T_D - T_A) - (C_D - C_A) \quad (4.8)$$

Trazendo a expressão (4.1) para o campo da regressão, a variável de interesse será o estimador de diferenças em diferenças (DID) α :

$$y_{it} = \beta_0 + \beta_1 G_{it} + \beta_2 T_t + \alpha G_{it} T_t + \varepsilon_{it} \quad (4.9)$$

As variáveis explicativas G_{it} e T_t são variáveis binárias, em que G_{it} se refere ao grupo de pesquisadores, e recebe o valor 1 (um) caso o pesquisador faça parte do grupo de tratamento e recebe zero se pertencer ao grupo de controle. A variável T_t é a variável indicativa do tempo analisado, recebendo o valor 1 (um) caso o período seja aquele em que ocorreu a obtenção da bolsa PQ e, zero caso se refira ao período anterior a essa obtenção e, ε_{it} é o termo de erro. O

valor estimado de β_1 corresponde às diferenças existentes entre o grupo de tratamento e de controle que não se devem ao efeito provocado pela obtenção da bolsa PQ, ou seja, capta as “diferenças naturais” entre esses dois grupos. Já o valor estimado de β_2 corresponde às alterações na carreira de todos os pesquisadores da amostra ocorridas entre o ano de obtenção de bolsa PQ e o ano em análise. Dessa forma, o parâmetro α que é o termo de interação entre as variáveis binárias G_{it} e T_t , mede a mudança ocorrida na carreira devido à obtenção de bolsa PQ, dado que essa política geraria impactos somente no grupo de tratamento.

4.3 Descrição das variáveis utilizadas no Modelo

Além do sistema de avaliação por pares, já discutido neste documento, a avaliação da ciência tem se baseado fortemente no uso de indicadores cientométricos em diferentes campos do conhecimento em todo o mundo, em virtude do crescimento no número de publicações em diferentes áreas do conhecimento, seja com a finalidade de se conhecer os pesquisadores e o avanço do conhecimento, ou para subsidiar políticas governamentais e de instituições voltadas para Ciência, Tecnologia e Inovação (CAREGNATO; VANZ, 2020; OLIVEIRA; GRACIO, 2011; STREHL, 2005).

Apesar de inúmeras controvérsias e ressalvas apontadas por pesquisadores, as análises bibliométricas têm se mostrado como procedimentos palpáveis e confiáveis. O que é preciso enfatizar, é que os indicadores clássicos de produção e citação, como o total de publicações, média de citações por documento e total de citações vem sendo amplamente utilizados e aceitos.

A variável dependente utilizada é IND_CC, que é o indicador do esforço da captação de recursos de pesquisa, gestão em CT&I e reconhecimento por pares. A *proxy* foi construída por intermédio da PCA, apresentado no Capítulo 3 e representa a capacidade do pesquisador em atrair recursos para pesquisa, por meio da coordenação de projetos de CT&I e de obter reconhecimento de seus pares por meio da participação em comitês e corpo editorial, bem como o recebimento de prêmios e títulos. Para a construção do IND_CC, como *proxy* de variável dependente, foram utilizadas as variáveis descritas no Quadro 3.7, vide Capítulo 3.

As variáveis explicativas usadas no trabalho estão apresentadas no Quadro 4.2.

Quadro 4.2 – Descrição das variáveis explicativas do modelo proposto

Variável	Descrição	Período	Fonte
Y17	Variável <i>Dummy</i> , indicativa do tempo analisado. Recebe 1 se o ano é 2017 e 0 para dos demais períodos	2013, 2014, 2017 e 2020	Elaboração própria
Y20	Variável <i>Dummy</i> , indicativa do tempo analisado. Recebe 1 se o ano é 2020 e 0 para dos demais períodos	2013, 2014, 2017 e 2020	Elaboração própria
TEMPO_DOU	Variável numérica. Tempo decorrido entre a obtenção do doutorado e a submissão da proposta.	2013, 2014, 2017 e 2020	Elaboração própria
GENERO	Variável categórica. Recebe Valor F, se do gênero Feminino e M, gênero masculino.	2013, 2014, 2017 e 2020	Plataforma Lattes, CNPq
REGIAO	Variável categórica. Recebe valor: NO, se da região Norte; NE- região nordeste; CO- região centro-oeste; S-região Sul e SE-região sudeste.	2013, 2014, 2017 e 2020	CNPq
IND_P	Variável Numérica. Indicador do esforço de publicação do pesquisador, reúne os indicadores de publicação em artigos, livros, capítulos de livros, organização de obra e trabalhos em eventos	2013, 2014, 2017 e 2020	Elaboração própria
IND_IMP	Variável Numérica. Indicador do impacto das publicações de cada pesquisador. Este indicador reúne as variáveis referentes ao impacto das publicações disponibilizados na planilha de produtividade.	2013, 2014, 2017 e 2020	Elaboração própria
IND_RH	Variável Numérica. Indicador do esforço e da capacidade do pesquisador em gerar recursos humanos por meio da orientação de trabalhos de conclusão de curso, mestrado, doutorado e supervisão de pós-doutorado.	2013, 2014, 2017 e 2020	Elaboração própria

Fonte: Elaboração própria.

A estimação de α foi feita por meio do modelo apresentado na equação 4.10, utilizando a variável IND_CC , como variável dependente, para período de 2013 a 2020.

$$IND_CC_{it} = \beta_0 + \beta_1 DTRAT_{it} + \beta_2 Yxx_t + \alpha DTRAT_{it} Yxx_t + \varepsilon_{it} \quad (4.10)$$

Onde Yxx_t corresponde à *dummy* indicativa do tempo analisado (2017 ou 2020). Adicionalmente foram inseridas variáveis de controle à equação (4.10) para estimar de forma mais eficiente o valor de α . As variáveis explicativas adicionais foram: $TEMPO_DOU$ (variável numérica, que representa o tempo de doutorado do pesquisador), $GENERO$ (variável binária tendo valor 1 para homens e valor igual a zero para mulheres), $REGIAO$ (região geográfica), IND_P (indicador criado para medir o esforço de publicação do pesquisador), IND_IMP (indicador criado para medir o impacto das publicações do pesquisadores) e IND_RH (indicador do esforço de formação de RH). Dessa forma, a equação a ser estimada toma a seguinte forma:

$$IND_CC_{it} = \beta_0 + \beta_1 DTRAT_{it} + \beta_2 Yxx_t + \alpha DTRAT_{it} Yxx_t + \beta_3 TEMPO_DOU_{it} + \beta_4 GENERO_{it} + \beta_5 REGIAO_{it} + \beta_6 IND_P_{it} + \beta_7 IND_IMP_{it} + \beta_8 IND_RH_{it} + \varepsilon_{it} \quad (4.11)$$

$$IND_CC_{i1} - IND_CC_{i0} = \alpha DTRAT_i + e_{i1} - e_{i0} \quad (4.12)$$

Para que o efeito seja não viesado é necessário assumir que a trajetória temporal dos grupos de controle e de tratamento seria a mesma caso não houvesse a intervenção, no caso, a obtenção de uma bolsa PQ. Dessa forma, o grupo de controle mostra-se um possível representante para o contrafactual do grupo de tratamento e as diferenças entre ambos os grupos refletiria o efeito da intervenção (obtenção da bolsa PQ). Tal hipótese é descrita como hipótese de tendências paralelas e pode ser representada pela expressão matemática (4.13):

$$E[e_{i1} - e_{i0} | DTRAT_i = 1] = E[e_{i1} - e_{i0} | DTRAT_i = 0] \quad (4.13)$$

Além das hipóteses de tendências paralelas, uma outra importante suposição do modelo DID é que não haja “choques” externos após a intervenção que afetem ambos os grupos de forma diferentes. Isso quer dizer, que se supõe que na ausência da intervenção (bolsa de produtividade PQ), ambos os grupos teriam evoluído de formas idênticas. Assim, a evolução da variável de impacto selecionada para ambos os grupos devem ser semelhantes.

4.4 Estatística descritiva das variáveis

De acordo com os resultados apresentados no Quadro 4.3, constata-se que para a variável que reflete o tempo de doutorado até a submissão de proposta de bolsa de produtividade (TEMPO_DOU), a média é 11,93 anos, com 75% (3º quartil) dos pesquisadores tendo até 16 anos de obtenção do doutoramento. Tendo em vista que o tempo mínimo exigido para obtenção de uma bolsa PQ-2 é de 3 anos (CNPQ, 2015), trata-se de um valor elevado, o que pode indicar que há um longo período entre o término do doutorado e a obtenção de uma bolsa PQ.

Os indicadores IND_P, IND_IMP, IND_RH e IND_CC foram obtidos através de técnica PCA e escalonadas para um valor entre 0 e 100, conforme explicado no Capítulo 3. Para a variável que reflete o esforço de publicação, IND_P, 75% dos pesquisadores estão no valor até 11,64, o que pode indicar que os pesquisadores em análise não possuem um volume muito grande de publicações. No entanto, quando observamos a variável que reflete o impacto das publicações IND_IMP, este mesmo percentual de 75% (3º quartil) estão próximos do valor máximo 100 (95,21), com média de 61,05; o que pode indicar que os pesquisadores se esforçam para publicar nos melhores periódicos. Assim, pelos dados apresentados, isso pode significar que os pesquisadores em análise, não procuram publicar em quantidade, mas em qualidade.

O indicador de esforço de formação de recursos humanos, IND_RH, aponta uma média de 29,02 com 3º quartil em 47,03, ou seja, 75% dos pesquisadores da amostra têm menos da metade do valor máximo nesse indicador.

O indicador de capital científico que reflete a capacidade do pesquisador em atrair recursos para pesquisa, por meio da coordenação de projetos de CT&I e de obter reconhecimento de seus pares por meio da participação em comitês e corpo editorial, bem como o recebimento de prêmios e títulos, tem uma média de 80,22, com 75% dos pesquisadores da amostra com até 98,10.

Quadro 4.3 - Estatística descritiva das variáveis do modelo proposto.

	TEMPO_DOU	IND_P	IND_IMP	IND_RH	IND_CC
Média	11,93	9,15	61,05	29,02	80,22
Mediana	11,00	5,68	87,63	16,90	93,70
Mínimo	3,00	0,00	0,00	0,00	0,00
Máximo	39,00	100,00	100,00	100,00	100,00
Desvio padrão	6,13	9,81	40,98	28,46	30,62
1º Quartil	7,00	3,52	9,89	6,92	83,34
3º Quartil	16,00	11,64	95,21	47,03	98,10

Elaboração própria a partir do Software R versão 4.1.2.

5. RESULTADOS OBTIDOS

Nesta seção são apresentados os resultados obtidos pelo estimador do Modelo de Diferenças-em-Diferenças (DID) aplicado aos dados em painel para estimação do efeito das bolsas de produtividade em pesquisa na carreira de pesquisadores brasileiros em dois períodos distintos, nos primeiros três anos após a obtenção da bolsa PQ (2014-2017) e seis anos após a obtenção da bolsa PQ (2014-2020).

A amostra é então composta por pesquisadores de 12 comitês assessores do CNPq. As variáveis foram coletadas para os anos de 2013, 2014, 2017 e 2020. Tem-se, dessa forma, um painel de dimensões $N = 1351$ e $T = 4$. Uma vez que os dados de alguns pesquisadores não estão presentes para todos os anos, o painel é dito ser não-balanceado.

5.1 Análise dos Grupos de Tratamento e Controle

Visando avaliar se a amostra é condizente com a metodologia DID, é fundamental a comparação do indicador IND_CC (indicador da obtenção de capital científico) entre os grupos de controle e de tratamento durante o período avaliado, por meio do teste de médias, o *Teste-t de Student*, o qual é apresentado na Tabela 5.1. Este teste mostra a diferença da média do indicador IND_CC do grupo de tratamento em relação ao grupo de controle, explicitado pelo p-valor da estatística bicaudal. Conforme Tabela 5.1, há uma diferença estatisticamente significativa para a amostra total antes de controlar pelas covariadas.

A Tabela 5.1 mostra o resultado do teste de médias para indicador de obtenção de capital científico (IND_CC), bem como o indicador do esforço de formação de recursos humanos (IND_RH) dos grupos de tratamento e controle, além das variáveis explicativas IND_P (indicador de esforço de publicação), IND_IMP (indicador de esforço de impacto das publicações) e TEMPO_DOU (tempo entre a obtenção do título de doutorado e a submissão de proposta de bolsa de produtividade PQ). Para este teste, a hipótese nula H_0 : $Média_T - Média_C = 0$, p-valor $> 0,05$. A hipótese alternativa é H_1 : $Média_T - Média_C \neq 0$, p-valor $\leq 0,05$.

$$\text{T-test} \left\{ \begin{array}{l} H_0: Média_T - Média_C = 0; \text{ p-valor} > 0,05 \\ H_1: Média_T - Média_C \neq 0; \text{ p-valor} \leq 0,05 \end{array} \right.$$

O teste mostra a significância do p-valor para rejeição da hipótese nula de que as médias são iguais considerando o nível de confiança de 95%. O resultado do teste permite afirmar que não se deve rejeitar H_0 , ou seja, confirma-se a presença de tendências paralelas – o grupo de tratamento e o grupo de possuem indicadores médios de capital científico estatisticamente iguais.

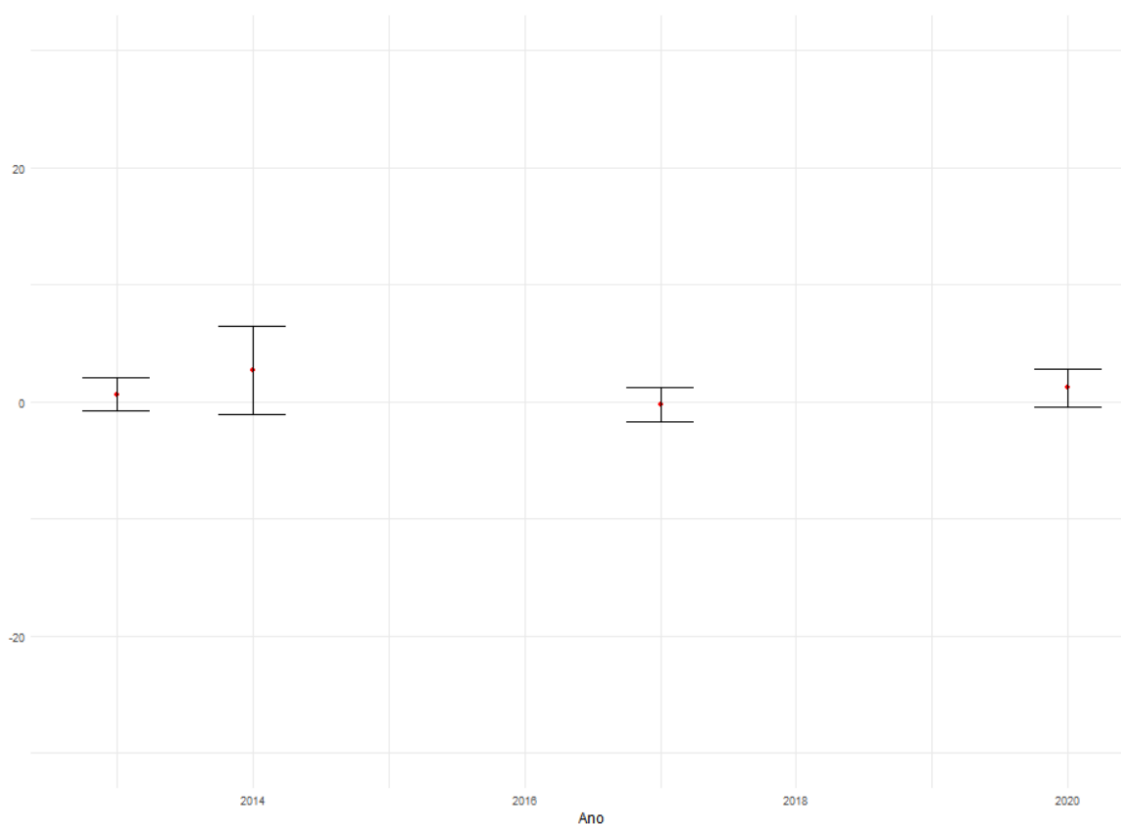
Tabela 5.1 – Teste-t de médias

	2013			2014			2017			2020		
	T	C	p-valor	T	C	p-valor	T	C	p-valor	T	C	p-valor
IND_CC	97,84	95,25	0,3771	83,91	81,23	0,1658	7,64	11,36	0,06321	94,34	90,77	0,09043
IND_RH	13,55	13,49	0,9691	18,27	18,60	0,8633	76,12	66,22	0,03999	73,32	69,17	0,1380
IND_P	4,14	5,03	0,0051	9,81	12,29	0,0031	16,81	22,71	0,0065	13,34	16,29	0,2261
IND_IMP	93,78	92,53	0,2119	5,78	7,22	0,1156	89,68	86,46	0,0724	22,84	25,16	0,4118
TEMPO_DOU	11,22	9,51	0,0050	11,79	10,51	0,05148	12,87	13,09	0,7691	16,38	15,21	0,1542

Fonte: CNPq, 2022. Elaboração própria.

De forma complementar, é possível notar a presença de tendências paralelas pela representação gráfica da diferença entre as médias de IND_CC das observações dos grupos de controle e tratamento e seus respectivos intervalos de confiança (95%), apresentada no Gráfico 5.1. O intuito deste gráfico é testar se os intervalos de confiança contêm o valor zero; caso isso ocorra, então, não há indício para rejeitar-se a hipótese nula de que as médias dos grupos de tratamento e controle sejam iguais a zero. Contudo, se o intervalo de confiança não contiver o valor zero, então é rejeitada a hipótese nula de que médias de IND_CC das observações dos grupos de controle e tratamento são iguais.

Gráfico 5.1 – Média da diferença entre o indicador de obtenção de capital científico (IND_CC) do grupo de controle e tratamento e seus respectivos limites inferiores e superiores dos intervalos de confiança (95%)

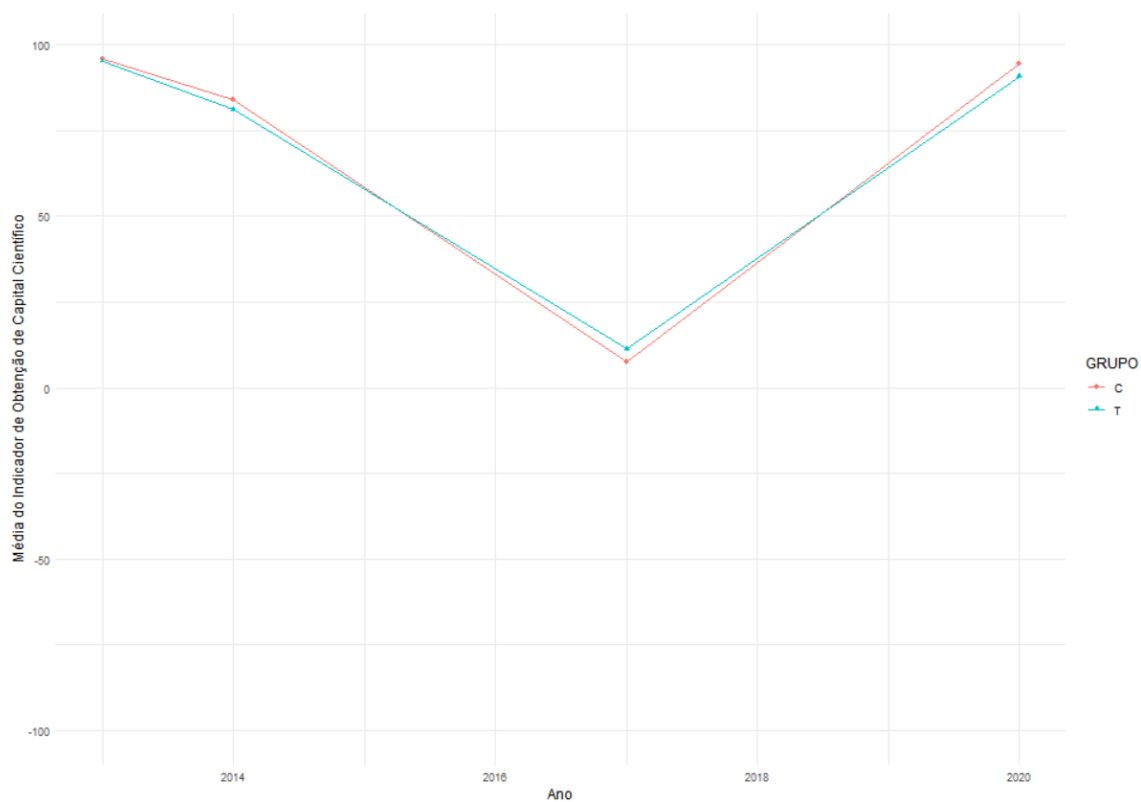


Fonte: CNPq, 2022. Elaborado própria a partir do Software R (versão 4.1.2).

Nota-se no Gráfico 5.1 que no ano de referência (2013), no ano em que estabeleceu a política (2014) e nos anos seguintes (2017 e 2020), a diferença entre as médias do indicador criado de obtenção de capital científico (IND_CC) das observações do grupo de tratamento e controle é estatisticamente igual a zero, ou seja, os intervalos de confiança contêm o valor zero.

No Gráficos 5.2 temos as médias do indicador IND_CC, para os anos em análise, dos grupos de tratamento e controle.

Gráfico 5.2 – Média do indicador de obtenção de capital científico (IND_CC) do grupo de controle e tratamento



Fonte: CNPq, 2022. Elaborado própria a partir do Software R (versão 4.1.2).

5.2 Resultados da Avaliação dos Efeitos da Bolsa de Produtividade PQ

Neste tópico são apresentados os resultados da aplicação dos diferentes modelos de dados em painel *pooled*, efeitos fixos e efeitos aleatórios para verificação dos efeitos da obtenção de uma bolsa de produtividade PQ na obtenção de capital científico, utilizando o estimador DID com dados em painel, para os anos de 2013 a 2020.

Os resultados das regressões para os anos de 2013 a 2017 e de 2013 a 2020 estão reportados nos Quadros 5.2 e 5.3, respectivamente; os quais mostram os valores estimados do parâmetro α da equação (4.11), vide capítulo 4:

$$IND_CC_{it} = \beta_0 + \beta_1 DTRAT_{it} + \beta_2 Yxx_{it} + \alpha DTRAT_{it} Yxx_{it} + \beta_3 TEMPO_DOU_{it} + \beta_4 GENERO_{it} + \beta_5 REGIAO_{it} + \beta_6 IND_P_{it} + \beta_7 IND_IMP_{it} + \beta_8 IND_RH_{it} + \varepsilon_{it}$$

Onde, Yxx_{it} corresponde à *dummy* indicativa do tempo analisado (2017 ou 2020), vide Quadro 4.2.

O parâmetro de interesse está no termo de interação $DTRAT_{it} Yxx_{it}$: α mede o efeito (a diferença da variação no tempo entre o grupo de tratamento e controle) no indicador de capital científico, IND_CC , que reflete a capacidade do pesquisador em atrair recursos para pesquisa, por meio da coordenação de projetos de CT&I e de obter reconhecimento de seus pares por meio da participação em comitês e corpo editorial, bem como o recebimento de prêmios e títulos. Este é o nosso estimador de diferenças em diferenças (DID), conforme apresentado no Capítulo 13 de Wooldridge (2016), desde que possamos presumir que tanto os pesquisadores com bolsa PQ quanto os pesquisadores sem bolsa, não tenham sido afetados, por outras razões.

Nos Quadros 5.1 e 5.2, os valores entre parênteses referem-se ao valor do teste t-Student e, estão destacados em negrito os coeficientes que foram significativos ao nível de 5% ou menos.

Inicialmente, foram feitas estimações da equação 4.11, utilizando 2017 em relação a 2013 (os primeiros anos de bolsa PQ-2) e os resultados estão apresentados no Quadro 5.1. Todos os modelos estimados foram estatisticamente significantes a 5%.

Na análise dos resultados dos modelos, no Quadro 5.1, pode-se verificar que:

- O efeito médio obtido ($DTRAT.Y17$) é positivo para todos os modelos. No modelo *pooled*, o efeito médio foi de 4,55, o que indica que a diferença da

variação no tempo entre o grupo de tratamento é de 4,55 sobre o capital científico em relação aos pesquisadores do grupo de controle. No modelo de efeitos fixos, esse efeito médio foi de 5,9 e para o modelo de efeitos aleatórios foi de 4,6 aproximadamente.

- A variável referente ao tempo de doutorado (TEMPO_DOU) apresenta sinal positivo para todos os modelos estimados, indicando que quanto maior o tempo de doutorado, maior o capital científico obtido. No entanto, esta variável foi estatisticamente significante apenas para o modelo de efeitos fixos (2,72).
- A variável referente ao gênero (GENERO) apresentou sinal negativo, estatisticamente significante, para os modelos de efeito *pooled* e efeitos aleatórios, -1,33 e -1,37, respectivamente. O que sugere que a variação no tempo entre o grupo de tratamento e controle, com respeito ao capital científico é menor para as mulheres que para os homens.
- A variável de região (REGIAO) apresentou sinal negativo para os modelos de efeito *pooled* e de efeitos aleatórios estatisticamente significante a 10%, mas apenas para a região Sul (a Região sudeste não aparece, pois é a região de referência), sugerindo que pesquisadores pertencentes a esta região tem uma variação menor no tempo com respeito à obtenção de capital científico comparativamente com pesquisadores da região sudeste (-1,4 aproximadamente para ambos os modelos).
- A variável referente ao esforço de publicação (IND_P) apresentou sinal positivo estatisticamente significativo a 5% para os modelos de efeito *pooled* e efeitos aleatórios (0,128 e 0,130 respectivamente). Para o modelo de efeitos fixos o valor foi negativo e igual a -0.132.
- A variável referente ao impacto das publicações (IND_IMP) apresentou sinal positivo estatisticamente significativo a 5% em todos os modelos analisados (0,131, 0,158 e 0,131 respectivamente), que é o resultado esperado, sugerindo que pesquisadores do grupo de controle tem uma variação positiva maior ao longo de tempo com respeito ao impacto das publicações.

Quadro 5.1 – Resultado da regressão para estimação do efeito de uma bolsa PQ na obtenção de capital científico (2017 em relação a 2013)

Variável dependente: IND_CC			
	<i>Pooled</i>	Efeitos Fixos	Efeitos Aleatórios
Variáveis Explicativas	(1)	(2)	(3)
DTRAT	-1.560** (0.779)		-1.547* (0.794)
Y17	-87.772*** (1.386)	-86.395*** (1.607)	-87.861*** (1.381)
TEMPO_DOU	0.013 (0.052)	2.722*** (0.281)	0.018 (0.053)
GENEROF	-1.333** (0.622)	5.746 (5.005)	-1.372** (0.634)
REGIAOCO	-0.120 (1.065)	-11.468 (9.940)	-0.104 (1.087)
REGIAONE	0.213 (0.843)	-18.326* (9.624)	0.199 (0.859)
REGIAONO	-1.640 (1.875)	-14.130 (15.628)	-1.640 (1.899)
REGIAOSU	-1.398* (0.798)	5.901 (7.885)	-1.400* (0.815)
IND_IMP	0.131*** (0.008)	0.158*** (0.009)	0.131*** (0.008)
IND_P	0.130*** (0.036)	-0.132** (0.063)	0.128*** (0.036)
IND_RH	-0.036** (0.014)	-0.182*** (0.025)	-0.036** (0.014)
DTRAT:Y17	4.555** (1.784)	5.904*** (1.814)	4.643*** (1.770)
Constant	84.740*** (1.052)		84.667*** (1.062)
Observações	1351	1351	1351
R²	0.880	0.928	0.883
R² Ajustado	0.879	0.861	0.882
Estatística F	816.565*** (df = 12; 1338)	816.718*** (df = 11; 695)	9,911.166***
Nota:	* p<0.1; ** p<0.05; *** p<0.01		

Elaboração própria a partir do Software R versão 4.1.2.

No Quadro 5.2 temos o efeito da obtenção de uma bolsa de produtividade PQ sobre o esforço de obtenção de capital científico sintetizado pela variável IND_CC, com respeito ao período de 2013-2020. Ou seja, foi utilizado o ano de 2020 em relação a 2013, o ano base. O resultado, apresentado no Quadro 5.2 apresentou sinal negativo para todos os modelos estimados.

Este resultado parece indicar, que em média o grupo de tratamento obteve uma variação inferior a observada para o grupo de controle, ou seja, a obtenção de uma bolsa PQ tem efeito médio negativo no indicador de obtenção de capital científico, quando se considera um período maior de tempo. No entanto, apenas o modelo de efeitos fixos apresentou relevância estatística (10%), com valor de -4,525. Ou seja, a diferença da variação no tempo entre pesquisadores do grupo de tratamento e controle, com respeito ao indicador IND_CC foi negativa e igual 4,525.

Quadro 5.2 – Resultado da regressão para estimação do efeito de uma bolsa PQ na obtenção de capital científico (2020 em relação a 2013)

Variável dependente: IND_CC			
	<i>Pooled</i>	Efeitos Fixos	Efeitos Aleatórios
Variáveis Explicativas	(1)	(2)	(3)
DTRAT	-4.742*** (1.253)		-4.371*** (1.415)
Y20	62.283*** (2.279)	125.259*** (3.143)	63.519*** (2.220)
TEMPO_DOU	0.298*** (0.086)	-15.000*** (0.629)	0.280*** (0.094)
GENEROF	-1.920* (1.027)	3.912 (6.571)	-1.713 (1.142)
REGIAOCO	-2.242 (1.758)	-11.989 (13.056)	-2.150 (1.962)
REGIAONE	0.384 (1.391)	-21.452* (12.671)	0.079 (1.543)
REGIAONO	2.426 (3.093)	-8.791 (20.534)	2.364 (3.335)
REGIAOSU	-1.285 (1.317)	0.299 (10.373)	-1.200 (1.481)
IND_IMP	0.009 (0.013)	-0.064*** (0.011)	0.007 (0.012)
IND_P	-0.537*** (0.056)	-0.317*** (0.081)	-0.633*** (0.059)
IND_RH	-0.928*** (0.023)	-0.384*** (0.032)	-0.930*** (0.023)
DTRAT:Y20	-0.896 (3.220)	-4.525* (2.641)	-1.447 (3.095)
Constant	103.149*** (1.643)		104.579*** (1.736)
Observações	1351	1351	1351
R ²	0.673	0.876	0.720
R ² Ajustado	0.670	0.759	0.718
Estatística F	229.110*** (df = 12; 1338)	446.318*** (df = 11; 695)	3,009.600**
<i>Nota:</i>	* p<0.1; ** p<0.05; *** p<0.01		

Elaboração própria a partir do Software R versão 4.1.2.

5.3 Discussões sobre os resultados obtidos

No Quadro 5.1, temos a verificação do efeito da bolsa PQ sobre a carreira dos pesquisadores ao final dos primeiros três anos de sua obtenção (2017), ou seja, ao final da vigência de uma bolsa PQ-2, que tem duração de 36 meses. No Quadro 5.2, temos a verificação do efeito da mesma bolsa PQ sobre a carreira dos pesquisadores nos três anos seguintes, após os três primeiros, ou seja, 2020. Assim, podemos verificar a diferença da variação no tempo entre tratamento e controle o efeito nos dois diferentes períodos.

Nos resultados apresentados no Quadro 5.1, referentes aos primeiros três primeiros anos de bolsa PQ, o efeito médio obtido (*DTRAT.Y17*) foi positivo para todos os modelos (em torno de 5 pontos) indicando que os pesquisadores com bolsa PQ (grupo de tratamento) tem um aumento sobre o capital científico em relação aos pesquisadores do grupo de controle.

Nos resultados apresentados no Quadro 5.2, era esperado um aumento no indicador *IND_CC* ao final de 2020 em relação a 2013, segundo a hipótese inicial, mas os resultados não foram nesse sentido. Para todos os modelos, o efeito médio obtido foi negativo e estatisticamente não significativo a 5%.

O grande objetivo desse estudo era demonstrar que a bolsa traz um ganho de capital científico aos pesquisadores e sua obtenção, traria assim um aumento nesse índice, tanto de 2017 em relação a 2013, quanto de 2020 em relação a 2013, comparando os grupos de tratamento e controle. Os resultados positivos e esperados, neste sentido, foram obtidos apenas quando estimamos o ano de 2017 em relação a 2013, conforme mostrado no Quadro 5.1.

Partindo deste ponto, a principal hipótese a ser levantada é justamente sobre os grupos de controle e de tratamento, os quais podem não ter sido bem escolhidos (se é que é possível tal escolha), pois conforme visto no Capítulo 3 há um dinamismo muito forte quanto à obtenção e perda destas bolsas nos grupos analisados, ao longo do período estudado.

Conforme visto no Capítulo 3, em 2013 foram selecionados todos os indivíduos que obtiveram parecer de recomendação dos comitês analisados, mas que não foram aprovados nas respectivas diretorias, ou seja, todos os indivíduos selecionados em 2013 não possuíam nenhuma bolsa PQ. O grupo de tratamento foi escolhido como sendo todos os pesquisadores que obtiveram bolsa PQ-2 em 2014 e que também haviam concorrido em 2013, mas não haviam obtido bolsa naquele ano. Como uma bolsa PQ-2 tem a duração de 36 meses, sabendo que os

pesquisadores que não possuem bolsa PQ continuam usualmente tentando, ano após ano, a proposta desta pesquisa era acompanhar a carreira destes pesquisadores ao longo do tempo. A ideia basicamente era a de que os pesquisadores que obtiveram bolsa PQ-2 em 2014, provavelmente submeteriam propostas (concorreriam novamente a uma bolsa PQ) em 2017 e 2020, e indivíduos que não conseguiram bolsa em 2014, iriam continuar tentando.

No entanto, durante a análise exploratória dos dados, verificou-se que dos 645 indivíduos selecionados em 2013, 371 submeteram propostas em 2014, 184 em 2017 e 151 ainda permaneciam em 2020. Em 2014 foi verificado que 98 dos 371 indivíduos foram contemplados com bolsa PQ-2. Esse grupo de 98 pesquisadores contemplados com bolsa PQ-2 em 2014, os quais também haviam enviado solicitação em 2013, mas não foram contemplados, foram caracterizados como o grupo de tratamento (T) e os demais como grupo de controle (C). Conforme podemos verificar no Quadro 3.12, o número de pesquisadores no grupo de controle e grupo de tratamento diminuiu ao longo dos anos, sendo a redução muito maior no grupo de controle (547 inicialmente e 95 ao final) e o que ocorre, é que estes pesquisadores ao não serem contemplados com bolsa PQ, continuam tentando nos anos subsequentes e muitos destes tentaram a obtenção de bolsa PQ em 2015 e 2016 e a obtiveram e da mesma forma em 2018 e 2019.

Fora este fato, ainda durante a análise exploratória de dados, verificou-se que há uma grande mobilidade entre os pesquisadores, pois dos 98 pesquisadores selecionados em 2014 para o grupo de tratamento, 12 progrediram de PQ-2 para PQ-1D em 2017 e assim, saem da amostra em 2020, pois a bolsa PQ-1D tem duração de 48 meses. Assim, dos 98 pesquisadores que foram contemplados com bolsa PQ-2 em 2014, 77 permaneceram na amostra em 2017. Ou seja, 21 pesquisadores saíram da amostra em 2017. Duas possibilidades podem ter ocorrido para isso: a) uma parte destes pesquisadores podem ser mulheres que pediram prorrogação parturiente (12 meses) por parto ocorrido no período da bolsa e assim, só vão submeter pedido novamente em 2018 (não analisado na pesquisa); ou b) se trata de uma questão de esquecimento de submissão, que é relativamente comum. Ainda com respeito ao grupo de tratamento, constatou-se que de 2017 para 2020, esse grupo foi reduzido de 77 para 56 pesquisadores. Excetuando os 12 que subiram de nível em 2017, o que pode ter ocorrido com os outros? A diferença mais uma vez, pode estar ligada às duas possibilidades já citadas anteriormente (prorrogação parturiente ou esquecimento).

Quanto ao grupo de controle, verificou-se que durante a análise exploratória de dados, que em 2017, todos os 107 indivíduos do grupo de controle foram contemplados com bolsa PQ. Sendo assim, em 2020, quase todos os pesquisadores entre grupo de controle e grupo de tratamento, tiveram bolsa em um dos períodos ou nos dois. Ainda assim, dos 107 indivíduos do grupo de controle que foram contemplados com bolsa PQ em 2017, 95 enviaram solicitação de bolsa em 2020, ou seja, 12 pesquisadores deixaram de submeter proposta (pode ter sido por prorrogação parturiente ou esquecimento). Destes 95, pertencentes ao grupo de controle, que haviam sido contemplados com bolsa PQ-2 em 2017 e submeteram novamente em 2020, 47 perderam a bolsa em 2020 e 48 continuaram com bolsa PQ-2. No entanto, isso não significa que 49% dos pesquisadores, pertencentes ao grupo de controle perderam a bolsa PQ de um julgamento para outro, já que foi verificado que não se tratava dos mesmos pesquisadores.

Essa mobilidade constatada tanto no grupo de tratamento quanto no grupo de controle necessitaria um estudo adicional envolvendo bolsistas de nível 1 e de nível 2, o que fica como sugestão para estudos futuros.

Do que foi relatado até aqui é preciso salientar, que dentro deste estudo, o ideal seria que o grupo de controle permanecesse sem bolsa ao longo do período analisado, ou seja sem “tratamento”. De forma que não foi possível isolar, para todo o período estudado, um grupo que não tivesse obtido tratamento. Isso só foi possível para o primeiro período estudado. Ainda assim, os resultados obtidos em 2020, são coerentes; assim como os obtidos em 2017. Em 2017, há uma clara distinção entre os grupos de tratamento e controle, o que não ocorreu em 2020, pois há indivíduos de ambos os grupos com bolsa PQ. Lembrando que em 2017, quase todos os indivíduos do grupo de controle obtiveram bolsa PQ.

Os resultados apresentados parecem sugerir que a obtenção de uma bolsa PQ na categoria 2 aumenta a obtenção de capital científico captado pelo indicador IND_CC, quando de fato podemos isolar indivíduos que possuem bolsa PQ e compará-los com aqueles que não possuem. Isso pode indicar, que de fato, há um esforço permanente por parte dos pesquisadores para obtenção destas bolsas de produtividade, assim como parece indicar um aspecto positivo de que o processo de julgamento efetuado pelos comitês assessores é de fato, um trabalho bem-feito, sendo necessário que os pesquisadores se mantenham sempre muito produtivos caso queiram manter a bolsa ou progredir de nível, ou poderão perder a bolsa no ciclo seguinte. Desta forma, tal efeito precisaria ser incorporado ao modelo estimado, o que também fica como sugestão para trabalhos futuros.

6. CONSIDERAÇÕES FINAIS

As bolsas de produtividade foram criadas em 1976, como forma de incentivar os pesquisadores brasileiros a aumentarem sua produtividade em C&T e em contrapartida permitir que o CNPq tivesse um corpo consistente de consultores para apoiá-lo em suas decisões e escolhas dos melhores projetos a serem financiados ou não. Desde sua criação, a bolsa PQ se mantém como um instrumento de diferenciação simbólica entre pares e com o tempo foi sofrendo uma considerável alteração de status, evoluindo como um “título” e praticamente se instituindo como uma carreira e ideal a ser alcançado na carreira dos pesquisadores no Brasil.

Por tudo isso, pode-se afirmar que as bolsas de produtividade PQ, permitem aos pesquisadores a acumulação das duas formas de capital científico abordados por Bourdieu e explanadas no Capítulo 2 desta dissertação, pois além de ser um apoio financeiro, a bolsa de produtividade é um sinal de reconhecimento pelos pares aos pesquisadores que possuam uma destacada produção científica e que permite aos cientistas a participação em uma série de processos fundamentais para a ciência brasileira, bem como abrem a possibilidade de captação de recursos diretos e indiretos, cujos efeitos se refletem de forma direta na estruturação das equipes e grupos de pesquisa.

Esta dissertação buscou avaliar os efeitos da obtenção de uma bolsa de Produtividade PQ na carreira dos pesquisadores brasileiros e para alcançar este objetivo, primeiro foi traçado um panorama sobre as bolsas PQ, seu histórico, sua importância e como se tornou o que é atualmente, bem como foi traçado o perfil médio atual do bolsista de produtividade. No estudo empírico foram analisados os dados de produtividade de pesquisadores de 12 comitês assessores referentes aos dois maiores comitês de cada uma das seis coordenações gerais do CNPq, os quais englobam todas as grandes áreas e cerca de 45% das bolsas de produtividade vigentes no CNPq. Esses dados foram reunidos em um painel de indicadores de produtividade científica dos pesquisadores presentes na amostra referentes aos anos de 2013, 2014, 2017 e 2020.

Para modelar e identificar o efeito de um conjunto de determinantes da obtenção de uma bolsa de produtividade PQ, foram estimados modelos de dados em painel com o estimador de diferenças em diferenças (DID). Para isso, inicialmente foram selecionados todos os pesquisadores que obtiveram parecer de recomendação em 2013, mas que não obtiveram bolsa PQ. Tal amostra foi dividida em dois grupos: tratamento e controle. Sendo o grupo de tratamento constituído pelos pesquisadores que obtiveram bolsa PQ em 2014. Ao final, obteve-se um painel desbalanceado de 1351 registros e quatro períodos (2013, 2014, 2017 e 2020).

O grande número de variáveis do painel foi agrupado em quatro dimensões (esforço de publicação; impacto das publicações, esforço de formação de recursos humanos e esforço de obtenção de capital científico) e para cada dimensão foi construído um indicador utilizando a técnica de PCA (*Principal Component Analysis*). Tais indicadores foram utilizados nos modelos de regressão de dados em painel com estimador DID, em três diferentes especificações: *pooled*, modelos com efeitos fixos e modelos com efeitos aleatórios.

O efeito médio (a diferença da variação no tempo entre tratamento e controle) encontrado para o ano de 2017 em relação a 2013, referente à obtenção de bolsa PQ foi positivo e estatisticamente significativo para os três modelos estudados. No entanto, o efeito médio encontrado para o ano de 2020 em relação a 2013, referente à obtenção de bolsa PQ foi negativo e estatisticamente (a 10%) significativo apenas para o modelo de efeitos fixos.

Os resultados obtidos apontam que a obtenção de uma bolsa PQ, de fato aumenta a obtenção de capital científico e, portanto, beneficia a carreira dos pesquisadores. No entanto, tal resultado só aparece quando podemos, de fato, isolar os indivíduos que não possuem bolsa PQ e compará-los com aqueles que possuem, o que foi possível para o ano de 2017, mas não de 2020. O estudo evidenciou que dentro do grupo de controle e do grupo de tratamento há bastante mobilidade com pesquisadores do grupo de controle recebendo bolsa em 2017 e/ou 2020, bem como pesquisadores saindo do grupo de tratamento. Desta forma, os resultados obtidos em 2020 ainda assim, são coerentes, pois de fato em 2020 não houve diferença estatística entre os grupos de tratamento e de controle.

Os resultados também parecem indicar duas coisas: 1) há, de fato, um enorme esforço por parte dos pesquisadores brasileiros não apenas para a obtenção destas bolsas de produtividade PQ, mas também para sua continuidade e para progressão de nível; e 2) o processo de julgamento efetuado pelos comitês assessores é de fato bem-feito, sendo necessário que os pesquisadores se mantenham continuamente produtivos e concorrendo com seus pares, caso queiram manter a bolsa ou subir de nível ou caso contrário, irão perde-la no ciclo seguinte. Ou seja, pelo menos para os pesquisadores nível 2, parece não haver “zona de conforto”. Contudo, tal efeito necessitaria de estudos adicionais, o que fica como sugestão para trabalhos futuros.

7. REFERÊNCIAS BIBLIOGRÁFICAS

1. ANDRADE, R.O. Desigualdade entre homens e mulheres marca a distribuição de bolsas de produtividade em pesquisa do CNPq. Pesquisa FAPESP. ed. 311, pg. 42-45, 2022.
2. AQUINO, Estela. M.L. Gênero e Ciência no Brasil: Contribuições para pensar a ação política na busca da equidade. Brasil/SPM. Pensando gênero e ciência. Encontro Nacional de Núcleos e Grupos de Pesquisas, 2006.
3. BALTAGI, B. H. Econometric analysis of panel data. John Wiley & Sons. 3rd ed. 2005.
4. BOURDIEU, PIERRE. O campo científico In: ORTIZ, Renato (org.). Pierre Bourdieu. Sociologia. São Paulo: Ática, 1983.
5. BOURDIEU, PIERRE. Os usos Sociais da Ciência: Por uma sociologia clínica do campo científico. São Paulo. Editora Unesp, 2004.
6. BRASIL. Lei Nº 1.310, de 15 de Janeiro de 1951. Cria o Conselho Nacional de Pesquisas, e dá outras providências. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/1950-1969/11310.htm. Acesso em 20/10/2021.
7. BREUSCH, T. S.; PAGAN, A.R. 1980. The Lagrange multiplier test and its applications to model specification in econometrics. The Review of Economic Studies 47 (1): 239–53.
8. CABRAL, M.V.F. Avaliação do impacto do INFOCRIM sobre as taxas de homicídios dos municípios paulistas: uma aplicação do método de diferenças em diferenças espacial. Tese de Doutorado defendida pela Faculdade de Economia. Universidade Federal de Juiz de Fora (UFJF), 2016.
9. CARD, D ; KRUEGER. A.B. Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania, The American Economic Review, 84(4), 772-793, 1994.
10. CAREGNATO, S. E.; VANZ, S. A. de S. Citações e indicadores de impacto na avaliação de revistas. INFORMAÇÃO & SOCIEDADE (UFPB. ONLINE), v. 30, p. 101-118, 2020.
11. CAVALCANTE, R. A. et al . Perfil dos pesquisadores da área de odontologia no Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Rev. bras. epidemiol., São Paulo , v. 11, n. 1, p. 106-113, mar. 2008 . Disponível em

- <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-790X2008000100010&lng=pt&nrm=iso>.
12. CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO (CNPq). Resolução Normativa 04/75. Consultoria Científica. Arquivo do Setor de Normas CNPq, 1975.
 13. _____. Resolução Normativa 01/76. Comitês Assessores (Especializados em Áreas do Conhecimento). Arquivo do Setor de Normas CNPq, 1976.
 14. _____. Resolução Normativa 012/88. Consultores "Ad Hoc". 1988. Disponível em <http://intranet.cnpq.br/normas/revogadas/normas-rn/rn-012-88.htm> Acesso em: 14/11/2021.
 15. CNPQ – Conselho Nacional de Desenvolvimento Científico e Tecnológico. Brasília: CNPq, 2021a. Disponível em: < <https://www.gov.br/cnpq/pt-br>>. Acesso em: 16/10/2021.
 16. _____. Chamada CNPq N ° 04/2021 - Bolsas de Produtividade em Pesquisa, 2021a. Disponível em: < http://memoria2.cnpq.br/web/guest/chamadas-publicas?p_p_id=resultadosportlet_WAR_resultadoscnpqportlet_INSTANCE_0ZaM&filtro=encerradas&startPage=1&buscaChamada=&ano= >. Acesso em: 16/10/2021.
 17. _____. Resolução Normativa RN 028/2015, 2015a. Disponível em: < http://memoria2.cnpq.br/view/-/journal_content/56_INSTANCE_0oED/10157/2958271?COMPANY_ID=10132 >. Acesso em: 16/10/2021.
 18. _____. Resolução Normativa 002/2015. Comitês de Assessoramento, Comitês Temáticos, Núcleos de Assessores em Tecnologia e Inovação, Núcleo de Assessores para Cooperação Internacional e Consultoria Ad Hoc, 2015b. Disponível em: <http://memoria2.cnpq.br/web/guest/view/-/journal_content/56_INSTANCE_0oED/10157/2409490/>. Acesso em: 16/10/2021.
 19. CORRAR, L. J.; PAULO, E.; DIAS FILHO, J. M. Análise multivariada. FIPECAFI: Atlas, 2007.
 20. CROISSANT, Y.; MILO, G. Panel data econometrics with R. John Wiley & Sons, 1st ed. 2019.
 21. FÁVERO, Luiz Paulo Lopes e BELFIORE, Patrícia Prado. Manual de análise de dados: estatística e modelagem multivariada com Excel, SPSS e Stata.. Rio de Janeiro: Elsevier, 2017.

22. GERTLER, Paul J. et al. Avaliação de Impacto na Prática. Banco Internacional para Reconstrução e Desenvolvimento/Banco Mundial. Washington: DC, 2018.
23. GHIRADINI, P. P. B. Regressão diferenças em diferenças: uma análise de fusões no setor hospitalar brasileiro. Dissertação (Mestrado em Economia) - Universidade de Brasília-UnB, 2015.
24. GREENE, W.H. Econometric analysis. Pearson. 8thed. 2018.
25. GUEDES, M. C.; AZEVEDO, N. ; FERREIRA, L. O. . A produtividade científica tem sexo? Um estudo sobre bolsistas de produtividade do CNPq. Cadernos Pagu (UNICAMP. Impresso), v.1, p. 367-399, 2015.
26. GUJARATI, D. N.; PORTER, D. C. Econometria Básica. 5ª Edição. Porto Alegre: AMGH, 2011.
27. HAIR Jr., J.F.; BLACK, W.C.; BABIN, B.J.; ANDERSON, R.E. & TATHAM, R.L. Análise multivariada de dados. 6.ed. Porto Alegre, Bookman, 2009. 688p.
28. HAUSMAN, J. A. 1978. Specification tests in econometrics. *Econometrica: Journal of the econometric society*, 1251–71.
29. HEISS,F. Using R for Introductory Econometrics. Createspace Independent Publishing Platform, 2016.
30. KLEIBER, C.; ZEILEIS, A. Applied Econometrics with R (Use R).Springer. 2011
31. LATTIN, J.M; CARROL, J.D; GREEN, P.E. Análise de dados multivariados. São Paulo, Cengage Learning, 2011.
32. LEITE, A.C.F. Olhares sobre Avaliação por Pares: Institucionalização e Limites da Avaliação por pares no CNPq (1975-2016). Tese (Doutorado em Educação em Ciências Química da Vida e Saúde (Ufsm - Furg)) - Universidade Federal do Rio Grande do Sul, 2018.
33. LEITE, A.C.F; ROCHA NETO, Ivan. Perfil dos Bolsistas de Produtividade em Pesquisa do CNPq em Educação. *Revista Brasileira de Ensino Superior, Passo Fundo*, v. 3, n. 4, p. 97-112, dez. 2017. ISSN 2447-3944. Disponível em: <<https://seer.imed.edu.br/index.php/REBES/article/view/2350/1539>>. Acesso em: 17 nov. 2018. doi:<https://doi.org/10.18256/2447-3944.2017.v3i4.2350>.
34. MACEDO, L. L. D. ; ROSÁRIO, FRANCISCO JOSÉ PEIXOTO . Influência das políticas industriais de inovação no comportamento inovativo da indústria nacional no período de 2003-2014. *NEXOS ECONÔMICOS*, v. 14, p. 21-45, 2020.
35. MALBOUISSON, C.; TIRYAKI, G. F. Econometria na Prática. Rio de Janeiro: Alta BooksEditora, 2017.

36. MELO, H.; LASTRES, H. M. M. . Ciência e Tecnologia numa perspectiva de gênero: o caso do CNPq. In: Lucy Woellner dos Santos; Elisa Y. Ichikawa; Doralice de F. Cargano. (Org.). Ciência Tecnologia e Gênero - desvelando o feminino na construção do conhecimento. Londrina: Instituto Agrônômico do Paraná - IAPAR, 2006, v., p. 129-160.
37. MINGOTI, S. A. Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada. Belo Horizonte: Editora UFMG, 2005.
38. NEVES, M. A. S. P. ; ANTUNES, R. C. ; RIBEIRO JUNIOR, E. S. ; GORGATI, C. Q. Evolução das bolsas de produtividade em pesquisa e dos editais universais do CNPq no Programa Básico de Zootecnia: 2002 a 2006. Anais da 44.^a Reunião Anual da Sociedade Brasileira de zootecnia. Viçosa/MG: Sociedade Brasileira de Zootecnia, v. 36. p. 369-376, 2007.
39. OLIVEIRA, A.; MELO, M.; RODRIGUES, Q. B. ; PEQUENO, Mayres . Gênero e desigualdade na academia brasileira: uma análise a partir dos bolsistas de produtividade em pesquisa do CNPq. Revista Configurações, v. 27, p. 74-93, 2021.
40. NURFALAH, A; ADIWIJAYA, K; ARDIYANTI, A. Cancer detection based on microarray data classification using PCA and modified back propagation. Far East Journal of Electronics and Communications, v. 16, p. 269-281, 2016.
41. OLIVEIRA, E. F. T.; GRACIO, M. C. C. . Indicadores bibliométricos em Ciência da Informação: análise dos pesquisadores mais produtivos no tema estudos métricos na base de dados Scopus. Perspectivas em Ciência da Informação (Impresso), v. 16, p. 16-28, 2011.
42. SHINN, T.; RAGOUE, P. Controvérsias sobre a ciência: Por uma sociologia transversalista da atividade científica. Editora 34. São Paulo, 2008.
43. STREHL, L.. O Fator de Impacto do ISI e a Avaliação da Produção Científica: aspectos conceituais e metodológicos. Ciência da Informação (Impresso), Brasília, v. 34, n.1, p. 19-27, 2005.
44. TEIXEIRA, F. O.; LAZARETTI, L. R. ; MARION FILHO, P. J. . Fatores Associados a Produtividade do Trabalho e a Transição Demográfica dos Estados Brasileiros. Revista de Desenvolvimento Econômico (Impresso), v. 3, p. 172-198, 2020.
45. WICKHAM G; GROLEMUND, G. R for Data Science. 1^a Edição. Editora: O'Reilly, 2017.
46. WOOLDRIDGE, J. Introdução à Econometria. 3^a Edição. Cengage Learning, 2017.

ANEXO A

PLANILHAS DE PRODUTIVIDADE: VARIÁVEIS UTILIZADAS ATÉ 2013

Descrição das colunas da planilha INDICADORES	
TERMINO_PRIMEIRO_GD	Data da primeira bolsa PQ no CNPq
DATA_PRIMEIRO_PQ	Ano que terminou o primeiro doutorado
LIDER_GRUPO_PESQUISA	Líder de grupos de pesquisa do CNPq (Apenas se for o 1º Líder)
POSSUI_PDE_PREVIO	Treinamento prévio de pós-doutorado no exterior (Verifica a Formação do Currículo ou se teve PDE pago na Folha de pagamento do CNPq)
INDICE_H_ISI_INFORMADO	Índice H informado ISI
MEDIA_H_ISI_INF_NIVEL	Número médio de H em cada nível de bolsa - ISI Informado (Média calculada daqueles que tem H maior que 0)
INDICE_H_ISI_CALCULADO	Índice H Calculado ISI
MEDIA_H_ISI_CAL_NIVEL	Número médio de H em cada nível de bolsa - ISI Calculado (Média calculada daqueles que tem H maior que 0)
INDICE_H_SCOPUS_INFORMADO	Índice H informado Scopus
MEDIA_H_SCOPUS_INF_NIVEL	Número médio de H em cada nível de bolsa - SCOPUS Informado (Média calculada daqueles que tem H maior que 0)
INDICE_H_SCOPUS_CALCULADO	Índice H Calculado Scopus
MEDIA_H_SCOPUS_CAL_NIVEL	Número médio de H em cada nível de bolsa - SCOPUS Calculado (Média calculada daqueles que tem H maior que 0)
CITACOES_ISI_INFORMADO	Total de Citações informado ISI
MEDIA_CIT_ISI_INF_NIVEL	Número médio de citações em cada nível de bolsa - ISI Informado (Média calculada daqueles que tem Citações maior que 0)
CITACOES_ISI_CALCULADO	Total de Citações Calculado ISI
MEDIA_CIT_ISI_CAL_NIVEL	Número médio de citações em cada nível de bolsa - ISI Calculado (Média calculada daqueles que tem Citações maior que 0)
CITACOES_SCOPUS_INFORMADO	Total de Citações informado Scopus
MEDIA_CIT_SCOPUS_INF_NIVEL	Número médio de citações em cada nível de bolsa - SCOPUS Informado (Média calculada daqueles que tem Citações maior que 0)
CITACOES_SCOPUS_CALCULADO	Total de Citações Calculado Scopus
MEDIA_CIT_SCOPUS_CAL_NIVEL	Número médio de citações em cada nível de bolsa - SCOPUS Calculado (Média calculada daqueles que tem Citações maior que 0)
PRODUCOES_EDUC_POP_CT	Nº de produções para educação e popularização de CeT
MEMBRO_CORPO_EDITORIAL	Nº de periódicos onde é membro de corpo editorial (Atual)
ARTIGOS_TOTAL	Nº de artigos publicados em periódicos (Planilha com 5 ou 10 anos)
ART_JCR_MAIOR_IGUAL_6	Nº de artigos publicados em periódicos JCR ≥ 6 (Planilha com 5 ou 10 anos)
ART_JCR_5_E_MEIO_ATE_6	Nº de artigos publicados em periódicos JCR entre 5,5 e 6 (Planilha com 5 ou 10 anos)
ART_JCR_5_ATE_5_E_MEIO	Nº de artigos publicados em periódicos JCR entre 5 e 5,5 (Planilha com 5 ou 10 anos)

ART_JCR_4_E_MEIO_ATE_5	Nº de artigos publicados em periodicos JCR entre 4,5 e 5 (Planilha com 5 ou 10 anos)
ART_JCR_4_ATE_4_E_MEIO	Nº de artigos publicados em periodicos JCR entre 4 e 4,5 (Planilha com 5 ou 10 anos)
ART_JCR_3_E_MEIO_ATE_4	Nº de artigos publicados em periodicos JCR entre 3,5 e 4 (Planilha com 5 ou 10 anos)
ART_JCR_3_ATE_3_E_MEIO	Nº de artigos publicados em periodicos JCR entre 3 e 3,5 (Planilha com 5 ou 10 anos)
ART_JCR_2_E_MEIO_ATE_3	Nº de artigos publicados em periodicos JCR entre 2,5 e 3 (Planilha com 5 ou 10 anos)
ART_JCR_2_ATE_2_E_MEIO	Nº de artigos publicados em periodicos JCR entre 2 e 2,5 (Planilha com 5 ou 10 anos)
ART_JCR_1_E_MEIO_ATE_2	Nº de artigos publicados em periodicos JCR entre 1,5 e 2 (Planilha com 5 ou 10 anos)
ART_JCR_1_ATE_1_E_MEIO	Nº de artigos publicados em periodicos JCR entre 1 e 1,5 (Planilha com 5 ou 10 anos)
ART_JCR_MEIO_ATE_1	Nº de artigos publicados em periodicos JCR entre 0,5 e 1 (Planilha com 5 ou 10 anos)
ART_JCR_ATE_MEIO	Nº de artigos publicados em periodicos JCR entre 0 e 0,5 (Planilha com 5 ou 10 anos)
ART_JCR_NULO	Nº de artigos publicados em periodicos JCR não informado (Planilha com 5 ou 10 anos)
ART_MENOR_IGUAL_10_AUTORES	Nº de artigos publicados em periodicos com 10 ou menos autores (Planilha com 5 ou 10 anos)
ART_ENTRE_11_19_AUTORES	Nº de artigos publicados em periodicos com 11 a 19 autores (Planilha com 5 ou 10 anos)
ART_MAIOR_IGUAL_20_AUTORES	Nº de artigos publicados em periodicos com 20 ou mais autores (Planilha com 5 ou 10 anos)
ART_PRIMEIRO_AUTOR	Nº de artigos publicados em periodicos como primeiro autor (Planilha com 5 ou 10 anos)
ART_ULTIMO_AUTOR	Nº de artigos completos publicados como último autor - somente se tiver mais de 1 autor (Planilha com 5 ou 10 anos)
ART_JCR_1_3_PRIM_OU_ULT_AUT	Nº de artigos publicados em periodicos - JCR maior 1 e menor 3 informando entre eles quanto o solicitante é primeiro ou último autor (Planilha com 5 ou 10 anos)
ART_JCR_1_3_SEG_OU_PENULT_AUT	Nº de artigos publicados em periodicos - JCR maior 1 e menor 3 informando entre eles quanto o solicitante é segundo ou penúltimo autor (Planilha com 5 ou 10 anos)
ART_JCR_MAIOR_3_PRI_OU_ULT_AUT	Nº de artigos publicados em periodicos - JCR maior ou igual a 3 informando entre eles quanto o solicitante é primeiro ou último autor (Planilha com 5 ou 10 anos)
ART_JCR_MAIOR_3_SEG_OU_PEN_AUT	Nº de artigos publicados em periodicos - JCR maior ou igual a 3 informando entre eles quanto o solicitante é segundo ou penúltimo autor (Planilha com 5 ou 10 anos)
MEDIA_ARTIGOS_PUB_NIVEL	Número médio de publicações em cada nível de bolsa
SOMATORIO_GERAL_JCR	Somatório de JCR artigos completos publicados (Planilha com 5 ou 10 anos)
MEDIA_GERAL_JCR	JCR médio de artigos publicados em periodicos (Planilha com 5 ou 10 anos)
MEDIA_AUTORES_JCR_MAIOR_1	Nº médio de autores em artigos publicados em periodicos com JCR maior que 1 (Planilha com 5 ou 10 anos)

ORIENT_IC_CONCLUIDOS	Nº de orientações em iniciação científica concluídos (Planilha com 5 ou 10 anos)
ORIENT_GM_CONCLUIDOS	Nº de orientações de mestrado concluídas (Planilha com 5 ou 10 anos)
ORIENT_GD_CONCLUIDOS	Nº de orientações de doutorados concluídas (Planilha com 5 ou 10 anos)
SUPERV_PD_CONCLUIDOS	Nº de supervisão de pós doutorados concluídas (Planilha com 5 ou 10 anos)
ORIENT_IC_ANDAMENTO	Nº de orientações em iniciação científica em andamento (Planilha com 5 ou 10 anos)
ORIENT_GM_ANDAMENTO	Nº de orientações de mestrado em andamento (Planilha com 5 ou 10 anos)
ORIENT_GD_ANDAMENTO	Nº de orientações de doutorados em andamento (Planilha com 5 ou 10 anos)
SUPERV_PD_ANDAMENTO	Nº de supervisão de pós doutorados em andamento (Planilha com 5 ou 10 anos)
ORIENT_IC_CONCLUIDOS_TOT	Nº de orientações em iniciação científica concluídos - total do pesquisador
ORIENT_GM_CONCLUIDOS_TOT	Nº de orientações de mestrado concluídas - total do pesquisador
ORIENT_GD_CONCLUIDOS_TOT	Nº de orientações de doutorados concluídas - total do pesquisador
SUPERV_PD_CONCLUIDOS_TOT	Nº de supervisão de pós doutorados concluídas - total do pesquisador
MEDIA_ORIENT_GD_NIVEL	Número médio de orientações e co-orientações de Doutorado em cada nível de bolsa
MEDIA_ORIENT_GM_NIVEL	Número médio de orientações e co-orientações de Mestrado em cada nível de bolsa
TRAB_PUBLICADOS_EVENTOS	Nº de trabalhos completos publicados em eventos (Planilha com 5 ou 10 anos)
LIVROS	Nº de livros (Planilha com 5 ou 10 anos)
CAPITULOS_LIVROS	Nº de capítulos de livros publicados (Planilha com 5 ou 10 anos)
ORGANIZACAO_OBRA	Nº de obras organizadas (Planilha com 5 ou 10 anos)
PATENTES	Nº de patentes registradas (Planilha com 5 ou 10 anos)
EVENTOS_ORGANIZADOR	Nº de eventos nos quais é organizador
EVENTOS_PALESTRANTE_CONVIDADO	Nº de eventos onde é participante convidado (Conferencista ou Simposista)
PARTICIP_COMITE_ASSESS	Nº de comitês de assessoramento onde participa
INOVACOES_PROJETOS_EMPRESAS	Número de inovações por meio de projetos com empresas
PREMIOS_E_TITULOS	Número de prêmios e títulos honoríficos recebidos
PROJETOS_CERTIFICADOS	Número de projetos certificados no Currículo Lattes
DATA_INICIO_NIVEL_ATUAL	Data de início no nível atual da bolsa
CURRICULO	Currículo do bolsista
NIVEL_ATUAL_PQ	Nível Atual da bolsa PQ

ANEXO B

PLANILHAS DE PRODUTIVIDADE: VARIÁVEIS UTILIZADAS EM 2014 E 2015

Descrição das colunas da planilha INDICADORES	
TERMINO_PRIMEIRO_GD	Ano que terminou o primeiro doutorado
DATA_PRIMEIRO_PQ	Data da primeira bolsa PQ no CNPq
LIDER_GRUPO_PESQUISA	Líder de grupos de pesquisa do CNPq (Apenas se for o 1º Líder)
POSSUI_PDE_PREVIO	Treinamento prévio de pós-doutorado no exterior (Verifica a Formação do Currículo ou se teve PDE pago na Folha de pagamento do CNPq)
INDICE_H_ISI_INFORMADO	Índice H informado ISI
MEDIA_H_ISI_INF_NIVEL	Número médio de H em cada nível de bolsa - ISI Informado (Média calculada daqueles que tem H maior que 0)
INDICE_H_ISI_CALCULADO	Índice H Calculado ISI
MEDIA_H_ISI_CAL_NIVEL	Número médio de H em cada nível de bolsa - ISI Calculado (Média calculada daqueles que tem H maior que 0)
INDICE_H_SCOPUS_INFORMADO	Índice H informado Scopus
MEDIA_H_SCOPUS_INF_NIVEL	Número médio de H em cada nível de bolsa - SCOPUS Informado (Média calculada daqueles que tem H maior que 0)
INDICE_H_SCOPUS_CALCULADO	Índice H Calculado Scopus
MEDIA_H_SCOPUS_CAL_NIVEL	Número médio de H em cada nível de bolsa - SCOPUS Calculado (Média calculada daqueles que tem H maior que 0)
CITACOES_ISI_INFORMADO	Total de Citações informado ISI
MEDIA_CIT_ISI_INF_NIVEL	Número médio de citações em cada nível de bolsa - ISI Informado (Média calculada daqueles que tem Citações maior que 0)
CITACOES_ISI_CALCULADO	Total de Citações Calculado ISI
MEDIA_CIT_ISI_CAL_NIVEL	Número médio de citações em cada nível de bolsa - ISI Calculado (Média calculada daqueles que tem Citações maior que 0)
CITACOES_SCOPUS_INFORMADO	Total de Citações informado Scopus
MEDIA_CIT_SCOPUS_INF_NIVEL	Número médio de citações em cada nível de bolsa - SCOPUS Informado (Média calculada daqueles que tem Citações maior que 0)
CITACOES_SCOPUS_CALCULADO	Total de Citações Calculado Scopus
MEDIA_CIT_SCOPUS_CAL_NIVEL	Número médio de citações em cada nível de bolsa - SCOPUS Calculado (Média calculada daqueles que tem Citações maior que 0)
PRODUCOES_EDUC_POP_CT	Nº de produções para educação e popularização de CeT
MEMBRO_CORPO_EDITORIAL	Nº de periódicos onde é membro de corpo editorial (Atual)
ARTIGOS_TOTAL	Nº de artigos publicados em periódicos (Planilha com 5 ou 10 anos)
ART_JCR_MAIOR_IGUAL_6	Nº de artigos publicados em periódicos JCR >= 6 (Planilha com 5 ou 10 anos)
ART_JCR_5_E_MEIO_ATE_6	Nº de artigos publicados em periódicos JCR entre 5,5 e 6 (Planilha com 5 ou 10 anos)
ART_JCR_5_ATE_5_E_MEIO	Nº de artigos publicados em periódicos JCR entre 5 e 5,5 (Planilha com 5 ou 10 anos)
ART_JCR_4_E_MEIO_ATE_5	Nº de artigos publicados em periódicos JCR entre 4,5 e 5 (Planilha com 5 ou 10 anos)
ART_JCR_4_ATE_4_E_MEIO	Nº de artigos publicados em periódicos JCR entre 4 e 4,5 (Planilha com 5 ou 10 anos)
ART_JCR_3_E_MEIO_ATE_4	Nº de artigos publicados em periódicos JCR entre 3,5 e 4 (Planilha com 5 ou 10 anos)
ART_JCR_3_ATE_3_E_MEIO	Nº de artigos publicados em periódicos JCR entre 3 e 3,5 (Planilha com 5 ou 10 anos)
ART_JCR_2_E_MEIO_ATE_3	Nº de artigos publicados em periódicos JCR entre 2,5 e 3 (Planilha com 5 ou 10 anos)

ART_JCR_2_ATE_2_E_MEIO	Nº de artigos publicados em periodicos JCR entre 2 e 2,5 (Planilha com 5 ou 10 anos)
ART_JCR_1_E_MEIO_ATE_2	Nº de artigos publicados em periodicos JCR entre 1,5 e 2 (Planilha com 5 ou 10 anos)
ART_JCR_1_ATE_1_E_MEIO	Nº de artigos publicados em periodicos JCR entre 1 e 1,5 (Planilha com 5 ou 10 anos)
ART_JCR_MEIO_ATE_1	Nº de artigos publicados em periodicos JCR entre 0,5 e 1 (Planilha com 5 ou 10 anos)
ART_JCR_ATE_MEIO	Nº de artigos publicados em periodicos JCR entre 0 e 0,5 (Planilha com 5 ou 10 anos)
ART_JCR_NULO	Nº de artigos publicados em periodicos JCR não informado (Planilha com 5 ou 10 anos)
ART_MENOR_IGUAL_10_AUTORES	Nº de artigos publicados em periodicos com 10 ou menos autores (Planilha com 5 ou 10 anos)
ART_ENTRE_11_19_AUTORES	Nº de artigos publicados em periodicos com 11 a 19 autores (Planilha com 5 ou 10 anos)
ART_MAIOR_IGUAL_20_AUTORES	Nº de artigos publicados em periodicos com 20 ou mais autores (Planilha com 5 ou 10 anos)
ART_PRIMEIRO_AUTOR	Nº de artigos publicados em periodicos como primeiro autor (Planilha com 5 ou 10 anos)
ART_ULTIMO_AUTOR	Nº de artigos completos publicados como último autor - somente se tiver mais de 1 autor (Planilha com 5 ou 10 anos)
ART_JCR_1_3_PRIM_OU_ULT_AUT	Nº de artigos publicados em periodicos - JCR maior 1 e menor 3 informando entre eles quanto o solicitante é primeiro ou último autor (Planilha com 5 ou 10 anos)
ART_JCR_1_3_SEG_OU_PENULT_AUT	Nº de artigos publicados em periodicos - JCR maior 1 e menor 3 informando entre eles quanto o solicitante é segundo ou penúltimo autor (Planilha com 5 ou 10 anos)
ART_JCR_MAIOR_3_PRI_OU_ULT_AUT	Nº de artigos publicados em periodicos - JCR maior ou igual a 3 informando entre eles quanto o solicitante é primeiro ou último autor (Planilha com 5 ou 10 anos)
ART_JCR_MAIOR_3_SEG_OU_PEN_AUT	Nº de artigos publicados em periodicos - JCR maior ou igual a 3 informando entre eles quanto o solicitante é segundo ou penúltimo autor (Planilha com 5 ou 10 anos)
MEDIA_ARTIGOS_PUB_NIVEL	Número médio de publicações em cada nível de bolsa
SOMATORIO_GERAL_JCR	Somatório de JCR artigos completos publicados (Planilha com 5 ou 10 anos)
MEDIA_GERAL_JCR	JCR médio de artigos publicados em periodicos (Planilha com 5 ou 10 anos)
MEDIA_AUTORES_JCR_MAIOR_1	Nº médio de autores em artigos publicados em periodicos com JCR maior que 1 (Planilha com 5 ou 10 anos)
ORIENT_IC_CONCLUIDOS	Nº de orientações em iniciação científica concluídos (Planilha com 5 ou 10 anos)
ORIENT_GM_CONCLUIDOS	Nº de orientações de mestrado concluídas (Planilha com 5 ou 10 anos)
ORIENT_GD_CONCLUIDOS	Nº de orientações de doutorados concluídas (Planilha com 5 ou 10 anos)
SUPERV_PD_CONCLUIDOS	Nº de supervisão de pós doutorados concluídas (Planilha com 5 ou 10 anos)
ORIENT_IC_ANDAMENTO	Nº de orientações em iniciação científica em andamento (Planilha com 5 ou 10 anos)
ORIENT_GM_ANDAMENTO	Nº de orientações de mestrado em andamento (Planilha com 5 ou 10 anos)
ORIENT_GD_ANDAMENTO	Nº de orientações de doutorados em andamento (Planilha com 5 ou 10 anos)
SUPERV_PD_ANDAMENTO	Nº de supervisão de pós doutorados em andamento (Planilha com 5 ou 10 anos)
ORIENT_IC_CONCLUIDOS_TOT	Nº de orientações em iniciação científica concluídos - total do pesquisador
ORIENT_GM_CONCLUIDOS_TOT	Nº de orientações de mestrado concluídas - total do pesquisador
ORIENT_GD_CONCLUIDOS_TOT	Nº de orientações de doutorados concluídas - total do pesquisador
SUPERV_PD_CONCLUIDOS_TOT	Nº de supervisão de pós doutorados concluídas - total do pesquisador

MEDIA_ORIENT_GD_NIVEL	Número médio de orientações e co-orientações de Doutorado em cada nível de bolsa
MEDIA_ORIENT_GM_NIVEL	Número médio de orientações e co-orientações de Mestrado em cada nível de bolsa
TRAB_PUBLICADOS_EVENTOS	Nº de trabalhos completos publicados em eventos (Planilha com 5 ou 10 anos)
LIVROS	Nº de livros (Planilha com 5 ou 10 anos)
CAPITULOS_LIVROS	Nº de capítulos de livros publicados (Planilha com 5 ou 10 anos)
ORGANIZACAO_OBRA	Nº de obras organizadas (Planilha com 5 ou 10 anos)
PATENTES	Nº de patentes registradas (Planilha com 5 ou 10 anos)
EVENTOS_ORGANIZADOR	Nº de eventos nos quais é organizador
EVENTOS_PALESTRANTE_CONVIDADO	Nº de eventos onde é participante convidado (Conferencista ou Simposista)
PARTICIP_COMITE_ASSESS	Nº de comitês de assessoramento onde participa
INOVACOES_PROJETOS_EMPRESAS	Número de inovações por meio de projetos com empresas
PREMIOS_E_TITULOS	Número de prêmios e títulos honoríficos recebidos
PROJETOS_CERTIFICADOS	Número de projetos certificados no Currículo Lattes
DATA_INICIO_NIVEL_ATUAL	Data de início no nível atual da bolsa
CURRICULO	Currículo do bolsista
NIVEL_ATUAL_PQ	Nível Atual da bolsa PQ
CO_ORIENT_GM_CONCLUIDOS	Nº de co-orientações de mestrado concluídas (Planilha com 5 ou 10 anos)
CO_ORIENT_GD_CONCLUIDOS	Nº de co-orientações de doutorados concluídas (Planilha com 5 ou 10 anos)
CO_ORIENT_GM_ANDAMENTO	Nº de co-orientações de mestrado em andamento (Planilha com 5 ou 10 anos)
CO_ORIENT_GD_ANDAMENTO	Nº de co-orientações de doutorados em andamento (Planilha com 5 ou 10 anos)
CO_ORIENT_GM_CONCLUIDOS_TOT	Nº de co-orientações de mestrado concluídas - total do pesquisador
CO_ORIENT_GD_CONCLUIDOS_TOT	Nº de co-orientações de doutorados concluídas - total do pesquisador

ANEXO C

PLANILHAS DE PRODUTIVIDADE: VARIÁVEIS UTILIZADAS DE 2016 EM DIANTE

Descrição das colunas da planilha INDICADORES	
TERMINO_PRIMEIRO_GD	Ano que terminou o primeiro doutorado
DATA_PRIMEIRO_PQ	Data da primeira bolsa PQ no CNPq
LIDER_GRUPO_PESQUISA	Líder de grupos de pesquisa do CNPq (Apenas se for o 1º Líder)
POSSUI_PDE_PREVIO	Treinamento prévio de pós-doutorado no exterior (Verifica a Formação do Currículo ou se teve PDE pago na Folha de pagamento do CNPq)
INDICE_H_ISI_INFORMADO	Índice H informado ISI
MEDIA_H_ISI_INF_NIVEL	Número médio de H em cada nível de bolsa - ISI Informado (Média calculada daqueles que tem H maior que 0)
INDICE_H_ISI_CALCULADO	Índice H Calculado ISI
MEDIA_H_ISI_CAL_NIVEL	Número médio de H em cada nível de bolsa - ISI Calculado (Média calculada daqueles que tem H maior que 0)
INDICE_H_SCOPUS_INFORMADO	Índice H informado Scopus
MEDIA_H_SCOPUS_INF_NIVEL	Número médio de H em cada nível de bolsa - SCOPUS Informado (Média calculada daqueles que tem H maior que 0)
INDICE_H_SCOPUS_CALCULADO	Índice H Calculado Scopus
MEDIA_H_SCOPUS_CAL_NIVEL	Número médio de H em cada nível de bolsa - SCOPUS Calculado (Média calculada daqueles que tem H maior que 0)
CITACOES_ISI_INFORMADO	Total de Citações informado ISI
MEDIA_CIT_ISI_INF_NIVEL	Número médio de citações em cada nível de bolsa - ISI Informado (Média calculada daqueles que tem Citações maior que 0)
CITACOES_ISI_CALCULADO	Total de Citações Calculado ISI
MEDIA_CIT_ISI_CAL_NIVEL	Número médio de citações em cada nível de bolsa - ISI Calculado (Média calculada daqueles que tem Citações maior que 0)
CITACOES_SCOPUS_INFORMADO	Total de Citações informado Scopus
MEDIA_CIT_SCOPUS_INF_NIVEL	Número médio de citações em cada nível de bolsa - SCOPUS Informado (Média calculada daqueles que tem Citações maior que 0)
CITACOES_SCOPUS_CALCULADO	Total de Citações Calculado Scopus
MEDIA_CIT_SCOPUS_CAL_NIVEL	Número médio de citações em cada nível de bolsa - SCOPUS Calculado (Média calculada daqueles que tem Citações maior que 0)
PRODUCOES_EDUC_POP_CT	Nº de produções para educação e popularização de CeT
MEMBRO_CORPO_EDITORIAL	Nº de periódicos onde é membro de corpo editorial (Atual)
ARTIGOS_TOTAL	Nº de artigos publicados em periódicos (Planilha com 5 ou 10 anos)
ART_JCR_MAIOR_IGUAL_6	Nº de artigos publicados em periódicos JCR ≥ 6 (Planilha com 5 ou 10 anos)
ART_JCR_5_E_MEIO_ATE_6	Nº de artigos publicados em periódicos JCR entre 5,5 e 6 (Planilha com 5 ou 10 anos)
ART_JCR_5_ATE_5_E_MEIO	Nº de artigos publicados em periódicos JCR entre 5 e 5,5 (Planilha com 5 ou 10 anos)
ART_JCR_4_E_MEIO_ATE_5	Nº de artigos publicados em periódicos JCR entre 4,5 e 5 (Planilha com 5 ou 10 anos)
ART_JCR_4_ATE_4_E_MEIO	Nº de artigos publicados em periódicos JCR entre 4 e 4,5 (Planilha com 5 ou 10 anos)
ART_JCR_3_E_MEIO_ATE_4	Nº de artigos publicados em periódicos JCR entre 3,5 e 4 (Planilha com 5 ou 10 anos)
ART_JCR_3_ATE_3_E_MEIO	Nº de artigos publicados em periódicos JCR entre 3 e 3,5 (Planilha com 5 ou 10 anos)
ART_JCR_2_E_MEIO_ATE_3	Nº de artigos publicados em periódicos JCR entre 2,5 e 3 (Planilha com 5 ou 10 anos)
ART_JCR_2_ATE_2_E_MEIO	Nº de artigos publicados em periódicos JCR entre 2 e 2,5 (Planilha com 5 ou 10 anos)
ART_JCR_1_E_MEIO_ATE_2	Nº de artigos publicados em periódicos JCR entre 1,5 e 2 (Planilha com 5 ou 10 anos)

ART_JCR_1_ATE_1_E_MEIO	Nº de artigos publicados em periodicos JCR entre 1 e 1,5 (Planilha com 5 ou 10 anos)
ART_JCR_MEIO_ATE_1	Nº de artigos publicados em periodicos JCR entre 0,5 e 1 (Planilha com 5 ou 10 anos)
ART_JCR_ATE_MEIO	Nº de artigos publicados em periodicos JCR entre 0 e 0,5 (Planilha com 5 ou 10 anos)
ART_JCR_NULO	Nº de artigos publicados em periodicos JCR não informado (Planilha com 5 ou 10 anos)
ART_QUALIS_A1	Nº de artigos publicados em periodicos Qualis A1 (Planilha com 5 ou 10 anos)
ART_QUALIS_A2	Nº de artigos publicados em periodicos Qualis A2 (Planilha com 5 ou 10 anos)
ART_QUALIS_B1	Nº de artigos publicados em periodicos Qualis B1 (Planilha com 5 ou 10 anos)
ART_QUALIS_B2	Nº de artigos publicados em periodicos Qualis B2 (Planilha com 5 ou 10 anos)
ART_QUALIS_B3	Nº de artigos publicados em periodicos Qualis B3 (Planilha com 5 ou 10 anos)
ART_QUALIS_B4	Nº de artigos publicados em periodicos Qualis B4 (Planilha com 5 ou 10 anos)
ART_QUALIS_B5	Nº de artigos publicados em periodicos Qualis B5 (Planilha com 5 ou 10 anos)
ART_QUALIS_C	Nº de artigos publicados em periodicos Qualis C (Planilha com 5 ou 10 anos)
ART_QUALIS_NULO	Nº de artigos publicados em periodicos Qualis não informado (Planilha com 5 ou 10 anos)
ART_MENOR_IGUAL_10_AUTORES	Nº de artigos publicados em periodicos com 10 ou menos autores (Planilha com 5 ou 10 anos)
ART_ENTRE_11_19_AUTORES	Nº de artigos publicados em periodicos com 11 a 19 autores (Planilha com 5 ou 10 anos)
ART_MAIOR_IGUAL_20_AUTORES	Nº de artigos publicados em periodicos com 20 ou mais autores (Planilha com 5 ou 10 anos)
ART_PRIMEIRO_AUTOR	Nº de artigos publicados em periodicos como primeiro autor (Planilha com 5 ou 10 anos)
ART_ULTIMO_AUTOR	Nº de artigos completos publicados como último autor - somente se tiver mais de 1 autor (Planilha com 5 ou 10 anos)
ART_JCR_1_3_PRIM_OU_ULT_AUT	Nº de artigos publicados em periodicos - JCR maior 1 e menor 3 informando entre eles quanto o solicitante é primeiro ou último autor (Planilha com 5 ou 10 anos)
ART_JCR_1_3_SEG_OU_PENULT_AUT	Nº de artigos publicados em periodicos - JCR maior 1 e menor 3 informando entre eles quanto o solicitante é segundo ou penúltimo autor (Planilha com 5 ou 10 anos)
ART_JCR_MAIOR_3_PRI_OU_ULT_AUT	Nº de artigos publicados em periodicos - JCR maior ou igual a 3 informando entre eles quanto o solicitante é primeiro ou último autor (Planilha com 5 ou 10 anos)
ART_JCR_MAIOR_3_SEG_OU_PEN_AUT	Nº de artigos publicados em periodicos - JCR maior ou igual a 3 informando entre eles quanto o solicitante é segundo ou penúltimo autor (Planilha com 5 ou 10 anos)
MEDIA_ARTIGOS_PUB_NIVEL	Número médio de publicações em cada nível de bolsa
SOMATORIO_GERAL_JCR	Somatório de JCR artigos completos publicados (Planilha com 5 ou 10 anos)
MEDIA_GERAL_JCR	JCR médio de artigos publicados em periodicos (Planilha com 5 ou 10 anos)
MEDIA_AUTORES_JCR_MAIOR_1	Nº médio de autores em artigos publicados em periodicos com JCR maior que 1 (Planilha com 5 ou 10 anos)
ORIENT_IC_CONCLUIDOS	Nº de orientações em iniciação científica concluídos (Planilha com 5 ou 10 anos)
ORIENT_GM_CONCLUIDOS	Nº de orientações de mestrado concluídas (Planilha com 5 ou 10 anos)
ORIENT_GD_CONCLUIDOS	Nº de orientações de doutorados concluídas (Planilha com 5 ou 10 anos)
SUPERV_PD_CONCLUIDOS	Nº de supervisão de pós doutorados concluídas (Planilha com 5 ou 10 anos)
ORIENT_IC_ANDAMENTO	Nº de orientações em iniciação científica em andamento (Planilha com 5 ou 10 anos)
ORIENT_GM_ANDAMENTO	Nº de orientações de mestrado em andamento (Planilha com 5 ou 10 anos)

ORIENT_GD_ANDAMENTO	Nº de orientações de doutorados em andamento (Planilha com 5 ou 10 anos)
SUPERV_PD_ANDAMENTO	Nº de supervisão de pós doutorados em andamento (Planilha com 5 ou 10 anos)
ORIENT_IC_CONCLUIDOS_TOT	Nº de orientações em iniciação científica concluídos - total do pesquisador
ORIENT_GM_CONCLUIDOS_TOT	Nº de orientações de mestrado concluídas - total do pesquisador
ORIENT_GD_CONCLUIDOS_TOT	Nº de orientações de doutorados concluídas - total do pesquisador
SUPERV_PD_CONCLUIDOS_TOT	Nº de supervisão de pós doutorados concluídas - total do pesquisador
MEDIA_ORIENT_GD_NIVEL	Número médio de orientações e co-orientações de Doutorado em cada nível de bolsa
MEDIA_ORIENT_GM_NIVEL	Número médio de orientações e co-orientações de Mestrado em cada nível de bolsa
TRAB_PUBLICADOS_EVENTOS	Nº de trabalhos completos publicados em eventos (Planilha com 5 ou 10 anos)
LIVROS	Nº de livros (Planilha com 5 ou 10 anos)
CAPITULOS_LIVROS	Nº de capítulos de livros publicados (Planilha com 5 ou 10 anos)
ORGANIZACAO_OBRA	Nº de obras organizadas (Planilha com 5 ou 10 anos)
PATENTES	Nº de patentes registradas (Planilha com 5 ou 10 anos)
PATENTES_TOTAL	Nº total de patentes registradas no Currículo Lattes
EVENTOS_ORGANIZADOR	Nº de eventos nos quais é organizador
EVENTOS_PALESTRANTE_CONVIDADO	Nº de eventos onde é participante convidado (Conferencista ou Simposista)
PARTICIP_COMITE_ASSESS	Nº de comitês de assessoramento onde participa
INOVACOES_PROJETOS_EMPRESAS	Número de inovações por meio de projetos com empresas
PREMIOS_E_TITULOS	Número de prêmios e títulos honoríficos recebidos
PROJETOS_CERTIFICADOS	Número de projetos certificados no Currículo Lattes
DATA_INICIO_NIVEL_ATUAL	Data de início no nível atual da bolsa
CURRICULO	Currículo do bolsista
NIVEL_ATUAL_PQ	Nível Atual da bolsa PQ
CO_ORIENT_GM_CONCLUIDOS	Nº de co-orientações de mestrado concluídas (Planilha com 5 ou 10 anos)
CO_ORIENT_GD_CONCLUIDOS	Nº de co-orientações de doutorados concluídas (Planilha com 5 ou 10 anos)
CO_ORIENT_GM_ANDAMENTO	Nº de co-orientações de mestrado em andamento (Planilha com 5 ou 10 anos)
CO_ORIENT_GD_ANDAMENTO	Nº de co-orientações de doutorados em andamento (Planilha com 5 ou 10 anos)
CO_ORIENT_GM_CONCLUIDOS_TOT	Nº de co-orientações de mestrado concluídas - total do pesquisador
CO_ORIENT_GD_CONCLUIDOS_TOT	Nº de co-orientações de doutorados concluídas - total do pesquisador
ORIENT_TCC_CONCLUIDAS	Nº de orientações de TCC concluídas (Planilha com 5 ou 10 anos)
PARTICIP_BANCA_DOUT_CONC	Nº de participação em bancas de doutorado concluídas (Planilha com 5 ou 10 anos)
PARTICIP_BANCA_MESTR_CONC	Nº de participação em bancas de mestrado concluídas (Planilha com 5 ou 10 anos)
COORD_PROJETO_PESQUISA	Nº de coordenações de projetos de pesquisa (Planilha com 5 ou 10 anos)
MEMBRO_EQUIPE_PROJ_PESQ	Nº de membros de equipe de projeto de pesquisa (Planilha com 5 ou 10 anos)
COORD_PROJETO_DESENV	Nº de coordenações de projeto de desenvolvimento (Planilha com 5 ou 10 anos)
MEMBRO_EQUIPE_PROJ_DESE	Nº de membros de equipe de projeto de desenvolvimento (Planilha com 5 ou 10 anos)
COORD_PROJETO_EXTENSAO	Nº de coordenações de projeto de extensão (Planilha com 5 ou 10 anos)
MEMBRO_EQUIPE_PROJ_EXTENSAO	Nº de membros de equipe de projeto de extensão (Planilha com 5 ou 10 anos)

APÊNDICE A

CONSTRUÇÃO DO PAINEL DE DADOS

```
##=====##
##=====PACOTES E BIBLIOTECAS=====##
##=====##

if(!require(ResourceSelection)) install.packages("ResourceSelection")
library(ResourceSelection)

if(!require(pacman)) install.packages("pacman")
library(pacman)

if(!require(modEvA)) install.packages("modEvA")
library(modEvA)

pacman::p_load(dplyr, psych, car, MASS, DescTools, QuantPsyc, ggplot2)

#install.packages("gridExtra")
#install.packages("readxl")
#install.packages("tidyverse")
#install.packages("questionr")
#install.packages("cov")
#install.packages("smooth")
#install.packages("var")
#install.packages("pROC")
#install.packages("caret")
#install.packages("leaps")
#install.packages("rio")
#install.packages("haven")
#install.packages("tidyr")
library(tidyr)
library(gridExtra)
library(rio)
library(haven)
library(plyr)
library(lmtest)
library(sandwich)
```

```
library(Biobase)
library(olsrr)
library(mctest)
library(GGally)
library(nortest)
library(car)
library(carData)
library(mctest)
library(caret)
library(leaps)
library(questionr)
library(cov)
library(smooth)
library(var)
library(pROC)
library(readr)
library(readxl)
library(stargazer)
library(dplyr)
library(tidyverse)
library(psych)
library(ggplot2)
library(readxl)
library(FactoMineR)
library(factoextra)
library(Factoshiny)
library(pacman)
library(psych)
library(REdaS)
library(writexl)
```

```
#####
#DRIVE DOS ARQUIVOS DE DADOS#
setwd ("C:/Users/55619/Documents/DADOS_TRATADOS_2")
#####
```

```

#####CARREGAMENTO DE PLANILHAS#####
#A PLANILHA (PLAN_DADOS_BRUTOS_2013_RC_DF.xlsx) CONTÉM OS DADOS DE
PESQUISADORES QUE
#FORAM RECOMENDADOS PELO COMITE, MAS NÃO TIVERAM AS BOLSAS APROVADAS
#ESSES SÃO DOS DADOS ANTES DO "TRATAMENTO"
#RC - RECOMENDADO
#DF - DESFAVORAVEL
#REC - RECORTE +> APENAS OS COMITES ESTUDADOS
#EM 2014, SÃO CRUZADOS OS DADOS DESTES PESQUISADORES COM A PLANILHA DE 2013
#AQUELES QUE OBTIVERAM BOLSA PQ-2 SÃO MARCADOS COMO GRUPO DE TRATAMENTO ==1
#AQUELES QUE FORAM RECOMENDADOS, MAS, NÃO OBTIVERAM BOLSA, SÃO MARCADOS
COM ==0
# E ESTES ENTRAM NO GRUPO DE CONTROLE == 0
#####

#####DADOS PLANILHAS DE DELIBERAÇÃO FINAL#####
####DADOS PLANILHAS DE DADOS BRUTOS -> CONTEM NUMERO DE CPF PARA UNIR
#TODAS AS PLANILHAS
#####
delib.2013 <- read_xlsx("PQ_DADOS_BRUTOS_2013_RC_DF.xlsx")
delib.2014 <- read_xlsx("PQ_DADOS_BRUTOS_2014_REC.xlsx")
delib.2017 <- read_xlsx("PQ_DADOS_BRUTOS_2017_REC.xlsx")
delib.2020 <- read_xlsx("PQ_DADOS_BRUTOS_2020_REC.xlsx")
summary(delib.2017)

#####
#carregando dados de produtividade de 2013, 2014,, 2017 e 2020
#PLANILHAS DE INDICADORES RECORTADA PARA OS COMITÊS ESTUDADOS
#####
dados.2013 <- read_xlsx("PLAN_INDIC_2013.xlsx")
dados.2014 <- read_xlsx("PLAN_INDIC_2014.xlsx")
dados.2017 <- read_xlsx("PLAN_INDIC_2017.xlsx")
dados.2020 <- read_xlsx("PLAN_INDIC_2020.xlsx")

#retira as variáveis ANO, COMITE, NOME, AREA, DATA_PRIMEIRO_PQ, NIVEL_ATUAL_PQ
#das 4 planilhas de indicadores

```

```

dados.2013 <- dados.2013[ , -c(1,2,4,5, 8, 83)]
dados.2014 <- dados.2014[ , -c(1,2,4,5, 8, 83)]
dados.2017 <- dados.2017[ , -c(1,2,4,5, 8,93)]
dados.2020 <- dados.2020[ , -c(1,2,4,5, 8,93)]

```

```

##=====##
##=====JUNTANDO PLANILHAS=====##

```

```

####OBS: O inner join despreza os registros de ambos os data.frames onde as#
###chaves não coincidem. Mas, existem situações em que esse descarte de #####
##registro não é interessante. Nesses casos usamos full join.#####

```

```

##=====##
#####EMPILHANDO AS PLANILHAS#####

```

```

#2013
#2014
#2017
#2020

```

```

#####JUNTANDO AS PLANILHAS#####
# juntando as planilhas de deliberação e de indicadores de produtividade

```

```

TOT_2013 <- inner_join(delib.2013, dados.2013, by = c("PROCESSO" = "PROCESSO"))
TOT_2014 <- inner_join(delib.2014, dados.2014, by = c("PROCESSO" = "PROCESSO"))
TOT_2017 <- inner_join(delib.2017, dados.2017, by = c("PROCESSO" = "PROCESSO"))
TOT_2020 <- inner_join(delib.2020, dados.2020, by = c("PROCESSO" = "PROCESSO"))

```

```

##cruzando os dados das planilhas de 2013 e 2014
# o objetivo é selecionar os grupos de controle e tratamento
#selecionar em 2014 quem obteve bolsa PQ-2, dentre os selcionados de 2013
# quem obteve Bolsa PQ =>grupo de controle
#quem não obteve + grupo de tratamento

```

```

TOT_2013_2014 <- bind_rows(TOT_2013, TOT_2014)
TOT_2013_CPF <- TOT_2013[ , c(5)]
TOT_T_C <- inner_join(TOT_2013_2014, TOT_2013_CPF, by = c("CPF" = "CPF"))

```

```

#SALVANDO TOT_T_C COMO UMA PLANILHA EXCEL E CRIO A VARIÁVEL GRUPO
write_xlsx(TOT_T_C, "PAINEL_2013_2014.xlsx")

```

```
#trazendo de volta a planilha
DADOS_TOT_T_C <- read_xlsx("PAINEL_2013_2014.xlsx")

#UNINDO AS DUAS PLANILHAS RESTANTES 2017 E 2020
TOT_2017_2020 <- bind_rows(TOT_2017, TOT_2020)

#FAZENDO A INTERSECÇÃO ENTRE TOT_T_C E A PLANILHA RESULTANTE DA UNIAO DAS
PLANILHAS 2017 E 2020
colnames(DADOS_TOT_T_C)
TOT_T_C_2013_2020 <- inner_join(TOT_2017_2020, TOT_2013_CPF, by = c("CPF" = "CPF"))
DADOS_FINAL <- bind_rows(DADOS_TOT_T_C, TOT_T_C_2013_2020)

#salvando em excel a planilha resultante
write_xlsx(DADOS_FINAL, "PAINEL_COMPLETO_2013_2020.xlsx")
```

APÊNDICE B

CONSTRUÇÃO DOS INDICADORES DO PAINEL

#####Construção indicadores painel 2013-2020#####

#MARIA MARONY SOUSA FARIAS

#MESTRADO EM ANÁLISE E MONITORAMENTO DE POLITICAS PUBLICAS - ENAP - 2022

#=====

#install.packages("readxl")

#install.packages("grDevices")

#install.packages("xlsx")

#install.packages("writexl")

#install.packages("scales")

library(scales)

library(writexl)

library(xlsx)

library(tidyr)

library(gridExtra)

library(rio)

library(haven)

library(plyr)

library(lmtest)

library(sandwich)

library(Biobase)

library(olsrr)

library(mctest)

library(GGally)

library(nortest)

library(car)

library(carData)

library(mctest)

library(caret)

library(leaps)

library(questionr)


```

library(cov)
library(smooth)
library(var)
library(pROC)
library(readr)
library(readxl)
library(stargazer)
library(dplyr)
library(tidyverse)
library(psych)
library(readxl)
library(FactoMineR)
library(factoextra)
library(Factoshiny)
library(pacman)
library(psych)
library(REdaS)
library(grDevices)

```

```

#CONSTRUCAO DE INDICADORES COM DADOS OBTIDOS DAS PLANILHAS DE INDICADORES DE
#PRODUTIVIDADE EM C,T&I DE 2013,2014, 2017 E 2020

```

```

#####

```

```

#DRIVE DOS ARQUIVOS DE DADOS#

```

```

setwd ("C:/Users/55619/Documents/DADOS_TRATADOS_2")

```

```

#-----#

```

```

#=====DESCRICAÇÃO PARTE I - SELEÇÃO DAS VARIÁVEIS=====#

```

```

#

```

```

#1. ETAPA 1 - CARREGA DADOS DA BASE

```

```

#2. ETAPA 2 - SELEÇÃO DAS VARIÁVEIS DE INTERESSE

```

```

#3. ETAPA 3 - SEPARAR AS VARIÁVEIS OBTIDAS POR ANO

```

```

#4. ETAPA 4 - PADRONIZAÇÃO DOS DADOS DE CADA ANO SEPARADAMENTE

```

```

#-----#

```

```

#1. ETAPA 1 - CARREGA DADOS DA BASE

```

```

dados.painel <- read_xlsx("PAINEL_COMPLETO_2013_2020.xlsx")

```

```

#=====#

```

#2. ETAPA 2 - SELECAO DAS VARIÁVEIS DE INTERESSE

#somente dados de publicacao utilizados ateh 2016, exclui os artigos qualis

#que so passaram a ser utilizados em 2017 em diante

#28 variaveis

#SOMENTE DADOS QUE REFLITAM O ESFORCO DE PUBLICACAO

#=====

```
public.total <- dados.painel[ , c(1,46:69, 88:91)]
```

```
summary(public.total)
```

#reúne todas as variáveis referentes ao impacto da produção de C&T (7 variáveis):

#variavel 2 -> ANO

```
impacto <- dados.painel[ , c(1,30,34,38,42,71:73)]
```

```
summary(impacto)
```

#reúne todas as variáveis referentes ao esforço de formação de RH (12 variáveis):

```
rh <- dados.painel[ , c(1, 74:85)]
```

```
summary(rh)
```

#reúne todas as variáveis referentes ao capital científico (06 variáveis):

```
cc <- dados.painel[ , c(1, 45,93:98)]
```

#exclui dados de c,t&i não disponíveis em 2013

```
summary(cc)
```

#todos os vetores acima incluem a variável da coluna 1: ANO

#54 variáveis serao assim reduzidas a 04 indices

#=====

#3. ETAPA 3 - SEPARAR AS VARIÁVEIS OBTIDAS POR ANO

#=====

#selecionar ano de 2013

```
public.total.2013 <- public.total%>%
```

```
  group_by(ANO) %>%
```

```
  filter(ANO==2013)
```

```
impacto.2013 <- impacto %>%
```

```
  group_by(ANO) %>%
```

```
  filter(ANO==2013)
```

```
rh.2013 <- rh %>%
```

```
  group_by(ANO) %>%
```

```
filter(ANO==2013)
```

```
cc.2013 <- cc %>%  
  group_by(ANO) %>%  
  filter(ANO==2013)  
summary(cc.2013)
```

```
#seleccionar ano de 2014
```

```
public.total.2014 <- public.total%>%  
  group_by(ANO) %>%  
  filter(ANO==2014)
```

```
impacto.2014 <- impacto %>%  
  group_by(ANO) %>%  
  filter(ANO==2014)
```

```
rh.2014 <- rh %>%  
  group_by(ANO) %>%  
  filter(ANO==2014)
```

```
cc.2014 <- cc %>%  
  group_by(ANO) %>%  
  filter(ANO==2014)
```

```
#selecciona ano de 2017
```

```
public.total.2017 <- public.total%>%  
  group_by(ANO) %>%  
  filter(ANO==2017)
```

```
summary(public.total.2017)
```

```
impacto.2017 <- impacto %>%  
  group_by(ANO) %>%  
  filter(ANO==2017)
```

```
rh.2017 <- rh %>%  
  group_by(ANO) %>%  
  filter(ANO==2017)
```

```
cc.2017 <- cc %>%  
  group_by(ANO) %>%  
  filter(ANO==2017)
```

```
#selecciona ano de 2020
```

```
public.total.2020 <- public.total%>%  
  group_by(ANO) %>%  
  filter(ANO==2020)
```

```
summary(public.total.2020)
```

```
impacto.2020 <- impacto %>%  
  group_by(ANO) %>%  
  filter(ANO==2020)
```

```
rh.2020 <- rh%>%  
  group_by(ANO) %>%  
  filter(ANO==2020)
```

```
cc.2020 <- cc %>%  
  group_by(ANO) %>%  
  filter(ANO==2020)
```

#4. ETAPA 4 - PADRONIZAÇÃO DOS DADOS DE CADA ANO SEPARADAMENTE

```
#=====#  
#=====PADRONIZACAO DE DADOS=====#  
#=====#
```

#PADRONIZAR OS DADOS CONSISTE DE UMA TECNICA DE REDIMENSIONAMENTO DE DADOS,

#EM QUE OS VALORES SAO CENTRALIZADOS EM TORNO DE UMA MEDIA DE UM DESVIO PADRAO DA UNIDADE. ISSO SIGNIFICA QUE A MEDIA DO ATRIBUTO TORNA-SE ZERO

#E A DISTRIBUICAO PADRAO TEM UM DESVIO PADRAO DA UNIDADE

#EH SENSIVEL A OUTLIERS

$X' = \frac{X - u}{dp}$

#A funcao scale faz isso de uma unica vez no data frame.

#os dados serao filtrados por ano

#sao padronizados os dados referentes a cada ano

#os dados de PCA de cada ano sao calculados com os dados padronizados

#uma coluna de indice eh criada (uma nova variavel) no data frame

#obtido anteriormente, e ao final junta-se todos os data frames

```
#=====#
```

#2013

stand.public.total.2013 <- scale(public.total.2013[,-c(1)]) #retira a variavel ano

stand.impacto.2013 <- scale(impacto.2013[,-c(1)]) #retira a variavel ano

stand.rh.2013 <- scale(rh.2013[,-c(1)]) #retira a variavel ano

stand.cc.2013 <- scale(cc.2013[,-c(1)]) #retira a variavel ano

#2014

stand.public.total.2014 <- scale(public.total.2014[,-c(1)]) #retira a variavel ano

stand.impacto.2014 <- scale(impacto.2014[,-c(1)]) #retira a variavel ano

stand.rh.2014 <- scale(rh.2014[,-c(1)]) #retira a variavel ano

stand.cc.2014 <- scale(cc.2014[,-c(1)]) #retira a variavel ano

summary(stand.impacto.2014)

```

#2017
stand.public.total.2017 <- scale(public.total.2017[,-c(1)]) #retira a variavel ano
summary(stand.public.total.2017)
stand.impacto.2017 <- scale(impacto.2017[,-c(1)]) #retira a variavel ano
stand.rh.2017 <- scale(rh.2017[,-c(1)]) #retira a variavel ano
stand.cc.2017 <- scale(cc.2017[,-c(1)]) #retira a variavel ano
summary(stand.cc.2017)

```

```

#2020

```

```

stand.public.total.2020 <- scale(public.total.2020[,-c(1)]) #retira a variavel ano
stand.impacto.2020 <- scale(impacto.2020[,-c(1)]) #retira a variavel ano
stand.rh.2020 <- scale(rh.2020[,-c(1)]) #retira a variavel ano
stand.cc.2020 <- scale(cc.2020[,-c(1)]) #retira a variavel ano e membro de equipe de desenv
summary(stand.cc.2020)

```

```

#Para extrair os fatores, vamos utilizar o metodo dos componentes principais.

```

```

#=====
#=====DESCRICAÇÃO PARTE II - CONSTRUÇÃO DOS INDICADORES=====
#ETAPAS PARA CONSTRUÇÃO DOS INDICES IND_P, IND_IMP, IND_RH E IND_CC #

```

```

#1. ETAPA 1 - OBTENÇÃO DOS COMPONENTES PRINCIPAIS COM DADOS PADRONIZADOS
#2. ETAPA 2 - RETIRAR O COMPONENTE 1 DOS COMPONENTES OBTIDOS (PCA1) PARA CADA
# ANO SEPARADAMENTE
#3. ETAPA 3 - JUNTAR AS COMPONENTES OBTIDAS PARA CADA INDICE (ANO A ANO)
# NUM UNICO VETOR
#4. ETAPA 4 - TRANSFORMAR O VETOR OBTIDO NUM DATA FRAME DE X LINHAS POR 1
COLUNA
# QUE SERÁ UMA NOVA COLUNA NA BASE DE DADOS
#5. ETAPA 5 - INTRODUIR ESTE DATA FRAME NO ARQUIVO PRINCIPAL COMO UMA NOVA
# VARIÁVEL

```

```

#=====
#1. ETAPA 1 - OBTENÇÃO DOS COMPONENTES PRINCIPAIS COM DADOS PADRONIZADOS
#=====
#1.INDICADOR DE PUBLICAÇÃO TOTAL
#=====

```

#1.COMPONENTES PRINCIPAIS 2013 - INDICE DE PUBLICAÇÃO TOTAL

```
pca.pub.tot.2013 <- prcomp(stand.public.total.2013)
```

```
cp.pub.tot.2013 <- pca.pub.tot.2013$x
```

```
summary(pca.pub.tot.2013)
```

#1.COMPONENTES PRINCIPAIS 2014 - INDICE DE PUBLICAÇÃO TOTAL

```
pca.pub.tot.2014 <- prcomp(stand.public.total.2014)
```

```
cp.pub.tot.2014 <- pca.pub.tot.2014$x
```

```
summary(pca.pub.tot.2014)
```

#1.COMPONENTES PRINCIPAIS 2017 - INDICE DE PUBLICAÇÃO TOTAL

```
pca.pub.tot.2017 <- prcomp(stand.public.total.2017)
```

```
summary(pca.pub.tot.2017)
```

```
cp.pub.tot.2017 <- pca.pub.tot.2017$x
```

#1.COMPONENTES PRINCIPAIS 2020 - - INDICE DE PUBLICAÇÃO TOTAL

```
pca.pub.tot.2020 <- prcomp(stand.public.total.2020)
```

```
summary(pca.pub.tot.2020)
```

```
cp.pub.tot.2020 <- pca.pub.tot.2020$x
```

```
#=====
```

#3. INDICADOR DE IMPACTO

```
#=====
```

```
#stand.impacto.2013
```

```
#stand.impacto.2014
```

```
#stand.impacto.2017
```

```
#stand.impacto.2020
```

#3.COMPONENTES PRINCIPAIS 2014 - INDICE DE IMPACTO

```
pca.impacto.2013 <- prcomp(stand.impacto.2013)
```

```
cp.impacto.2013 <- pca.impacto.2013$x
```

```
summary(pca.impacto.2013)
```

#3.COMPONENTES PRINCIPAIS 2014 - INDICE DE IMPACTO

```
pca.impacto.2014 <- prcomp(stand.impacto.2014)
cp.impacto.2014 <- pca.impacto.2014$x
summary(pca.impacto.2014)
```

#3.COMPONENTES PRINCIPAIS 2017 - INDICE DE IMPACTO

```
pca.impacto.2017 <- prcomp(stand.impacto.2017)
cp.impacto.2017 <- pca.impacto.2017$x
summary(pca.impacto.2017)
```

#3.COMPONENTES PRINCIPAIS 2020 - INDICE DE IMPACTO

```
pca.impacto.2020 <- prcomp(stand.impacto.2020)
cp.impacto.2020 <- pca.impacto.2020$x
summary(pca.impacto.2020)
```

```
#=====
```

#4. INDICADOR DE ESFORÇO DE FORMAÇÃO DE RH

```
#=====
```

#4.COMPONENTES PRINCIPAIS 2014 - INDICE DE ESFORÇO DE FORMAÇÃO DE RH

```
pca.rh.2013 <- prcomp(stand.rh.2013)
cp.rh.2013 <- pca.rh.2013$x
summary(pca.rh.2013)
```

#4.COMPONENTES PRINCIPAIS 2014 - INDICE DE ESFORÇO DE FORMAÇÃO DE RH

```
pca.rh.2014 <- prcomp(stand.rh.2014)
cp.rh.2014 <- pca.rh.2014$x
summary(pca.rh.2014)
```


#4.COMPONENTES PRINCIPAIS 2017 - INDICE DE ESFORÇO DE FORMAÇÃO DE RH

```
pca.rh.2017 <- prcomp(stand.rh.2017)
cp.rh.2017 <- pca.rh.2017$x
summary(pca.rh.2017)
```

#4.COMPONENTES PRINCIPAIS 2020 - INDICE DE ESFORÇO DE FORMAÇÃO DE RH

```
pca.rh.2020 <- prcomp(stand.rh.2020)
cp.rh.2020 <- pca.rh.2020$x
summary(pca.rh.2020)
```

```
#=====
```

#5. INDICADOR DE ESFORÇO DE GESTÃO EM C,T&I

```
#=====
```

```
#stand.cc.2013
#stand.cc.2014
#stand.cc.2017
#stand.cc.2020
```

#5.COMPONENTES PRINCIPAIS 2013 - INDICE REFERENTE AO CAPITAL CIENTÍFICO OBTIDO

```
pca.cc.2013 <- prcomp(stand.cc.2013)
cp.cc.2013 <- pca.cc.2013$x
summary(pca.cc.2013)
```

#5.COMPONENTES PRINCIPAIS 2014 - INDICE REFERENTE AO CAPITAL CIENTÍFICO OBTIDO

```
pca.cc.2014 <- prcomp(stand.cc.2014)
cp.cc.2014 <- pca.cc.2014$x
summary(pca.cc.2014)
```

#5.COMPONENTES PRINCIPAIS 2017 - INDICEREFERENTE AO CAPITAL CIENTÍFICO OBTIDO

```
pca.cc.2017 <- prcomp(stand.cc.2017)
cp.cc.2017 <- pca.cc.2017$x
summary(pca.cc.2017)
```

#5.COMPONENTES PRINCIPAIS 2020 - INDICADOR REFERENTE AO CAPITAL CIENTÍFICO OBTIDO

```
pca.cc.2020 <- prcomp(stand.cc.2020)
cp.cc.2020 <- pca.cc.2020$x
summary(pca.cc.2020)
```

```
#=====
```

```
#2. ETAPA 2 - RETIRAR O COMPONENTE 1 DOS COMPONENTES OBTIDOS (PCA1) PARA CADA
```

```
# ANO SEPARADAMENTE
```

```
#=====
```

```
#1.CONSTRUCAO DO INDICADOR DE PUBLICAÇÃO TOTAL
```

```
#=====
```

```
INDPT = cp.pub.tot.2013[ , c(1)]
```

```
IND_PT1 = as.data.frame(INDPT)
```

```
plot(IND_PT1)
```

```
boxplot(IND_PT1)
```

```
INDPT = cp.pub.tot.2014[ , c(1)]
```

```
IND_PT2 = as.data.frame(INDPT)
```

```
plot(IND_PT2)
```

```
boxplot(IND_PT2)
```

```
INDPT = cp.pub.tot.2017[ , c(1)]
```

```
IND_PT3 = as.data.frame(INDPT)
```

```
plot(IND_PT3)
```

```
boxplot(IND_PT3)
```

```
INDPT = cp.pub.tot.2020[ , c(1)]
```

```
IND_PT4 = as.data.frame(INDPT)
```

```
plot(IND_PT4)
```

```
boxplot(IND_PT4)
```

```
#=====
```

```
#3.CONSTRUCAO DO INDICADOR DE IMPACTO
```

```
#=====
```

```
INDIMP = cp.impacto.2013[ , c(1)]
```

```
IND_IMP1 = as.data.frame(INDIMP)
```

```
plot(IND_IMP1)
```

```
boxplot(IND_IMP1)
```

```
INDIMP = cp.impacto.2014[ , c(1)]
```

```
IND_IMP2 = as.data.frame(INDIMP)
```

```
plot(IND_IMP2)
```

```
boxplot(IND_IMP2)
```

```
INDIMP = cp.impacto.2017[ , c(1)]
```

```
IND_IMP3 = as.data.frame(INDIMP)
```

```
plot(IND_IMP3)
```

```
boxplot(IND_IMP3)
```

```
INDIMP = cp.impacto.2020[ , c(1)]
```

```
IND_IMP4 = as.data.frame(INDIMP)
```

```
plot(IND_IMP4)
```

```
boxplot(IND_IMP4)
```

```
#=====
```

```
#4.CONSTRUCAO DO INDICADOR DE ESFORÇO DE FORMAÇÃO DE RH
```

```
#=====
```

```
#=====
```

```
INDRH = cp.rh.2013[ , c(1)]
```

```
IND_RH1 = as.data.frame(INDRH)
```

```
plot(IND_RH1)
```

```
boxplot(IND_RH1)
```

```
INDRH = cp.rh.2014[ , c(1)]
```

```
IND_RH2 = as.data.frame(INDRH)
plot(IND_RH2)
boxplot(IND_RH2)
```

```
INDRH = cp.rh.2017[ , c(1)]
IND_RH3 = as.data.frame(INDRH)
plot(IND_RH3)
boxplot(IND_RH3)
```

```
INDRH = cp.rh.2020[ , c(1)]
IND_RH4 = as.data.frame(INDRH)
plot(IND_RH4)
boxplot(IND_RH4)
```

```
#=====#
#5.CONSTRUCAO DO INDICADOR DE ESFORÇO DE GESTAO EM C,T&I
#=====#
```

```
INDCC = cp.cc.2013[ , c(1)]
IND_CC1 = as.data.frame(INDCC)
plot(IND_CC1)
boxplot(IND_CC1)
```

```
INDCC = cp.cc.2014[ , c(1)]
IND_CC2 = as.data.frame(INDCC)
plot(IND_CC2)
boxplot(IND_CC2)
```

```
INDCC = cp.cc.2017[ , c(1)]
IND_CC3 = as.data.frame(INDCC)
plot(IND_CC3)
boxplot(IND_CC3)
```

```
INDCC = cp.cc.2020[ , c(1)]
IND_CC4 = as.data.frame(INDCC)
plot(IND_CC4)
boxplot(IND_CC4)
```

```

#=====
#3. ETAPA 3 - JUNTAR AS COMPONENTES OBTIDAS PARA CADA INDICE (ANO A ANO)
# NUM UNICO VETOR
#=====juntando os indices=====

dados.2013 <- dados.painel %>% filter(ANO == 2013)
dados.2014 <- dados.painel %>% filter(ANO == 2014)
dados.2017 <- dados.painel %>% filter(ANO == 2017)
dados.2020 <- dados.painel %>% filter(ANO == 2020)

##IND_P
dados.2013$IND_P <- IND_PT1[,c(1)]
dados.2013$IND_P <- as.numeric(dados.2013$IND_P)

dados.2014$IND_P <- IND_PT2[,c(1)]
dados.2014$IND_P <- as.numeric(dados.2014$IND_P)

dados.2017$IND_P <- IND_PT3[,c(1)]
dados.2017$IND_P <- as.numeric(dados.2017$IND_P)

dados.2020$IND_P <- IND_PT4[,c(1)]
dados.2020$IND_P <- as.numeric(dados.2020$IND_P)

##IND_IMP
dados.2013$IND_IMP <- IND_IMP1[,c(1)]
dados.2013$IND_IMP <- as.numeric(dados.2013$IND_IMP)

dados.2014$IND_IMP <- IND_IMP2[,c(1)]
dados.2014$IND_IMP <- as.numeric(dados.2014$IND_IMP)

dados.2017$IND_IMP <- IND_IMP3[,c(1)]
dados.2017$IND_IMP <- as.numeric(dados.2017$IND_IMP)

dados.2020$IND_IMP <- IND_IMP4[,c(1)]
dados.2020$IND_IMP <- as.numeric(dados.2020$IND_IMP)

```

```
#IND_RH
```

```
dados.2013$IND_RH <- IND_RH1[,c(1)]
```

```
dados.2013$IND_RH <- as.numeric(dados.2013$IND_RH)
```

```
dados.2014$IND_RH <- IND_RH2[,c(1)]
```

```
dados.2014$IND_RH <- as.numeric(dados.2014$IND_RH)
```

```
dados.2017$IND_RH <- IND_RH3[,c(1)]
```

```
dados.2017$IND_RH <- as.numeric(dados.2017$IND_RH)
```

```
dados.2020$IND_RH <- IND_RH4[,c(1)]
```

```
dados.2020$IND_RH <- as.numeric(dados.2020$IND_RH)
```

```
#IND_CC
```

```
dados.2013$IND_CC <- IND_CC1[,c(1)]
```

```
dados.2013$IND_CC <- as.numeric(dados.2013$IND_CC)
```

```
dados.2014$IND_CC <- IND_CC2[,c(1)]
```

```
dados.2014$IND_CC <- as.numeric(dados.2014$IND_CC)
```

```
dados.2017$IND_CC <- IND_CC3[,c(1)]
```

```
dados.2017$IND_CC <- as.numeric(dados.2017$IND_CC)
```

```
dados.2020$IND_CC <- IND_CC4[,c(1)]
```

```
dados.2020$IND_CC <- as.numeric(dados.2020$IND_CC)
```

```
#REESCALONANDO OS DADOS DOS INDICES CRIADOS PARA FICAREM NUMA ESCALA ENTRE 0  
E 100
```

```
#ONDE O INDIVÍDUO COM MENOR INDICE TERÁ VALOR 0 NESSA NOVA ESCALA, E O  
INDIVÍDUO COM MAIOR INDICE TERÁ VALOR 100
```

```
#2013
```

```
dados.2013$IND_PN <- rescale(dados.2013$IND_P, to = c(0, 100))
```

```
summary(dados.2013$IND_PN)
```

```
boxplot(dados.2013$IND_PN)
```

```
dados.2013$IND_IMP <- rescale(dados.2013$IND_IMP, to = c(0, 100))
```

```
summary(dados.2013$IND_IMP_N)  
boxplot(dados.2013$IND_IMP_N)
```

```
dados.2013$IND_RHN <- rescale(dados.2013$IND_RH, to = c(0, 100))  
summary(dados.2013$IND_RHN)  
boxplot(dados.2013$IND_RHN)
```

```
dados.2013$IND_CCN <- rescale(dados.2013$IND_CC, to = c(0, 100))  
summary(dados.2013$IND_CCN)  
boxplot(dados.2013$IND_CCN)
```

```
#2014
```

```
dados.2014$IND_PN <- rescale(dados.2014$IND_P, to = c(0, 100))  
summary(dados.2014$IND_PN)  
boxplot(dados.2014$IND_PN)
```

```
dados.2014$IND_IMP_N <- rescale(dados.2014$IND_IMP, to = c(0, 100))  
summary(dados.2014$IND_IMP_N)  
boxplot(dados.2014$IND_IMP_N)
```

```
dados.2014$IND_RHN <- rescale(dados.2014$IND_RH, to = c(0, 100))  
summary(dados.2014$IND_RHN)  
boxplot(dados.2014$IND_RHN)
```

```
dados.2014$IND_CCN <- rescale(dados.2014$IND_CC, to = c(0, 100))  
summary(dados.2014$IND_CCN)  
boxplot(dados.2014$IND_CCN)
```

```
#2017
```

```
dados.2017$IND_PN <- rescale(dados.2017$IND_P, to = c(0, 100))  
summary(dados.2017$IND_PN)  
boxplot(dados.2017$IND_PN)
```

```
dados.2017$IND_IMP_N <- rescale(dados.2017$IND_IMP, to = c(0, 100))
```

```

summary(dados.2017$IND_IMP_N)
boxplot(dados.2017$IND_IMP_N)

dados.2017$IND_RHN <- rescale(dados.2017$IND_RH, to = c(0, 100))
summary(dados.2017$IND_RHN)
boxplot(dados.2017$IND_RHN)

dados.2017$IND_CCN <- rescale(dados.2017$IND_CC, to = c(0, 100))
summary(dados.2017$IND_CCN)
boxplot(dados.2017$IND_CCN)

#2020
dados.2020$IND_PN <- rescale(dados.2020$IND_P, to = c(0, 100))
summary(dados.2020$IND_PN)
boxplot(dados.2020$IND_PN)

dados.2020$IND_IMP_N <- rescale(dados.2020$IND_IMP, to = c(0, 100))
summary(dados.2020$IND_IMP_N)
boxplot(dados.2020$IND_IMP_N)
dados.2020$IND_RHN <- rescale(dados.2020$IND_RH, to = c(0, 100))
summary(dados.2020$IND_RHN)
boxplot(dados.2020$IND_RHN)
dados.2020$IND_CCN <- rescale(dados.2020$IND_CC, to = c(0, 100))
summary(dados.2020$IND_CCN)
boxplot(dados.2020$IND_CCN)

#=====#
#=====INTRODUZINDO OS INDICES CRIADOS NA BASE=====#
#=====#

dados.painel<- bind_rows(dados.2013, dados.2014, dados.2017, dados.2020)
summary(dados.painel)

#=====salvando em xlsx=====#

write_xlsx(dados.painel, "PAINEL_2013_2022_COM_INDICES.xlsx")

```


APÊNDICE C

ESTIMAÇÃO DOS MODELOS DE REGRESSÃO

```
#=====#
#=====ANALISE DADOS EM PAINEL 2013, 2014, 2017 E 2020 =====#
#MPAM 2022
#MARIA MARONY SOUSA FARIAS
#=====#
#=====CHAMANDO AS BIBLIOTECAS=====#
#install.packages("foreign")
#install.packages("plm")
#install.packages("pglm")
#install.packages("faraway")
library(tidyr)
library(gridExtra)
library(rio)
library(haven)
library(plyr)
library(lmtest)
library(sandwich)
library(Biobase)
library(olsrr)
library(mctest)
library(GGally)
library(nortest)
library(car)
library(carData)
library(mctest)
library(caret)
library(leaps)
library(questionr)
library(cov)
library(smooth)
library(var)
library(pROC)
library(readr)
```

```

library(readxl)
library(stargazer)
library(dplyr)
library(tidyverse)
library(psych)
library(readxl)
library(FactoMineR)
library(factoextra)
library(Factoshiny)
library(pacman)
library(psych)
library(REdaS)
library(grDevices)
library(foreign)#pacote de dados em painel
library(plm)#pacote de dados em painel
library(pglm)#pacote dados em painel logit
library(faraway)
#####
#=====DRIVE DOS ARQUIVOS DE DADOS=====#

setwd ("C:/Users/55619/Documents/DADOS_TRATADOS_2")

#####
#=====CARREGANDO A BASE DE DADOS=====#
#BASE DE DADOS CONSTRUÍDA COM DADOS DE BOLSISTAS PQ DE 2014, 2017 E 2020#
#CARREGA DADOS
dados.painel <- read_xlsx("PAINEL_2013_2020_COM_INDICES.xlsx")
glimpse(dados.painel)
summary(dados.painel)

dados.painel$GRUPO= as.factor(dados.painel$GRUPO)
dados.painel$REGIAO= as.factor(dados.painel$REGIAO)
dados.painel$GENERO= as.factor(dados.painel$GENERO)
dados.painel$DTRAT= as.factor(dados.painel$DTRAT)
dados.painel$Y14= as.factor(dados.painel$Y14)
dados.painel$Y17= as.factor(dados.painel$Y17)

```

```
dados.painel$Y20= as.factor(dados.painel$Y20)
```

```
dados.painel$REGIAO <- relevel(dados.painel$REGIAO, "SE")
```

```
dados.painel$GENERO <- relevel(dados.painel$GENERO, "M")
```

```
#colocando os dados em forma de painel
```

```
painel<- pdata.frame(dados.painel, index=c("CPF", "ANO") )
```

```
#Observação: neste banco as variáveis IND_CCN, IND_PN, IND_IMP_N e IND_RHN foram criados com variáveis que existiam no ano base 2013 e foram escalonadas para ficarem entre os valores 0 e 100
```

```
# conforme explicado no documento de dissertação.
```

```
##+++++++Testes de especificação: multicolinearidade+++++++
```

```
##=====TESTE DE MULTICOLINEARIDADE=====
```

```
#criando um vetor somente com as variáveis explicativas
```

```
explicativas <- dados.painel[ , c(25,130:133)]
```

```
cor(explicativas)
```

```
summary(explicativas)
```

```
pairs(explicativas)
```

```
pairs.panels(explicativas)
```

```
vif(explicativas)
```

```
##=====
```

```
##Determinantes de obtenção de capital científico
```

```
#Modelo Pooled#
```

```
reg.ad.ep <- plm(IND_CCN ~ DTRAT
```

```
+ Y17
```

```
+ Y20
```

```
+ TEMPO_DOU
```

```
+ GENERO
```

```
+ REGIAO
```

```
+ IND_IMP_N
```

```
+ IND_PN
```

```
+ IND_RHN
, data = painel, model="pooling")
summary(reg.ad.ep)
```

```
## Modelo Efeitos Fixos
```

```
reg.ad.ef <- plm(IND_CCN ~ DTRAT
+ Y17
+ Y20
+ TEMPO_DOU
+ GENERO
+ REGIAO
+ IND_IMPNI
+ IND_PNI
+ IND_RHN
, data = painel, model="within")
summary(reg.ad.ef)
```

```
## Modelo Efeitos Aleatórios
```

```
reg.ad.ea <- plm(IND_CCN ~ DTRAT
+ Y17
+ Y20
+ TEMPO_DOU
+ GENERO
+ REGIAO
+ IND_IMPNI
+ IND_PNI
+ IND_RHN
, data = painel, model="random")
summary(reg.ad.ea)
```

```
stargazer(reg.ad.ep, reg.ad.ef, reg.ad.ea, type = "text", title = "Determinantes na obtenção de Capital científico",
out = "REG_DET_CC1.html")
```

Comparação e escolha dos modelos

1) Modelo Pooled x Modelo de Efeitos Fixos

#H0: há igualdade nos interceptos e nas inclinações para todos os indivíduos, caracterizando o modelo de dados agrupados (pooled).

pFtest(reg.ad.ep, reg.ad.ef)

2) Modelo Pooled x Modelo de Efeitos Aleatórios

#H0: hipótese nula implica que o modelo de dados agrupados (pooled) é preferível

plmtest(reg.ad.ep, type="bp")

3) Modelo Efeitos Fixos x Modelo de Efeitos Aleatórios

#H0: Efeitos Fixos é o mais adequado

phtest(reg.ad.ea, reg.ad.ef)

PARTE 2 - ESTIMADOR DE DIFERENÇAS EM DIFERENÇAS - (DID)

#Modelo Pooled

```
reg.ep.cc1 <- plm(IND_CCN ~ DTRAT
```

```
  + Y17
```

```
  + DTRAT*Y17
```

```
  + TEMPO_DOU
```

```
  + GENERO
```

```
  + REGIAO
```

```
  + IND_IMP
```

```
  + IND_PN
```

```
  + IND_RHN
```

```

        ,data = painel, model="pooling")
summary(reg.ep.cc1)

#Modelo Efeitos Fixos

reg.ef.cc1 <- plm(IND_CCN ~ DTRAT
+ Y17
+ DTRAT*Y17
+ TEMPO_DOU
+ GENERO
+ REGIAO
+ IND_IMP
+ IND_PN
+ IND_RHN
, data = painel, model="within")
summary(reg.ef.cc1)

#Modelo Efeitos Aleatórios

reg.ea.cc1 <- plm(IND_CCN ~ DTRAT
+ Y17
+ DTRAT*Y17
+ TEMPO_DOU
+ GENERO
+ REGIAO
+ IND_IMP
+ IND_PN
+ IND_RHN
, data = painel, model="random")
summary(reg.ea.cc1)

stargazer(reg.ep.cc1, reg.ef.cc1, reg.ea.cc1, type = "text", title = "Resultados", out =
"REG_IND_CC_23_08.txt")

stargazer(reg.ep.cc1, reg.ef.cc1, reg.ea.cc1, type = "text", title = "Resultados", out =
"REG_IND_CC_23_08.html")

#utilizando variáveis com dados completos com relação a 2020

```

```
#Modelo Pooled
```

```
reg.ep.cc2 <- plm(IND_CCN ~ DTRAT  
  + Y20  
  + DTRAT*Y20  
  + TEMPO_DOU  
  + GENERO  
  + REGIAO  
  + IND_IMP  
  + IND_PN  
  + IND_RHN  
  ,data = painel, model="pooling")  
summary(reg.ep.cc2)
```

```
#Modelo Efeitos Fixos
```

```
reg.ef.cc2 <- plm(IND_CCN ~ DTRAT  
  + Y20  
  + DTRAT*Y20  
  + TEMPO_DOU  
  + GENERO  
  + REGIAO  
  + IND_IMP  
  + IND_PN  
  + IND_RHN  
  ,data = painel, model="within")  
summary(reg.ef.cc2)
```

```
#Modelo Efeitos Aleatórios
```

```

reg.ea.cc2 <- plm(IND_CCN ~ DTRAT
  + Y20
  + DTRAT*Y20
  + TEMPO_DOU
  + GENERO
  + REGIAO
  + IND_IMP
  + IND_PN
  + IND_RHN
  ,data = painel, model="random")

summary(reg.ea.cc2)

stargazer(reg.ep.cc2, reg.ef.cc2, reg.ea.cc2, type = "text", title = "Resultados", out =
"REG_IND_CC2_23_08.html")

#      Comparação e escolha dos modelos

# 1)      Modelo Pooled x Modelo de Efeitos Fixos

#H0: há igualdade nos interceptos e nas inclinações para todos os indivíduos, caracterizando o modelo de dados
agrupados (pooled).

pFtest(reg.ep.cc1,reg.ef.cc1)
pFtest(reg.ep.cc2,reg.ef.cc2)

# 2) Modelo Pooled x Modelo de Efeitos Aleatórios

#H0: hipótese nula implica que o modelo de dados agrupados (pooled) é preferível

plmtest(reg.ep.cc1, type="bp")
plmtest(reg.ea.cc2, type="bp")

# 3) Modelo Efeitos Fixos x Modelo de Efeitos Aleatórios

#H0: Efeitos Fixos é o mais adequado

phtest(reg.ep.cc1,reg.ef.cc1)
phtest(reg.ep.cc2,reg.ef.cc2)

# Alguns testes para os modelos

# Teste de Normalidade dos resíduos Shapiro-Milk e Kolmogorov-Smirnov

#HO: normalidade nos resíduos da regressão

shapiro.test(reg.ep.cc1$residuals)
ks.test(reg.ep.cc1$residuals,pnorm)

```



```
# Teste de Heterocedasticidade
#H0: não há heterocedasticidade
bptest(reg.ep.cc1)
bptest(reg.ea.cc1)
bptest(reg.ef.cc1)

#Gráfico dos resíduos
residuos <- reg.ep.cc1$residuals
qqPlot(residuos)
```