

PROJEÇÕES EXPONENCIAIS DA CIÊNCIA BRASILEIRA: MODELOS E ANÁLISES QUANTITATIVAS DA PRODUÇÃO CIENTÍFICA NACIONAL PUBLICADA NOS ÚLTIMOS 30 ANOS

EXPONENTIAL PROJECTIONS OF BRAZILIAN SCIENCE: MODELS AND QUANTITATIVE ANALYSIS OF NATIONAL SCIENTIFIC PRODUCTION PUBLISHED IN THE LAST 30 YEARS

Dandara Souza Araújo Nascimento^a

Roney Fraga Souza^b

Jaim José da Silva Junior^c

Lucas Rodrigo da Silva^d

RESUMO

Introdução: O estudo examina o crescimento da produção científica brasileira no período 1990 – 2017, com base em amostra composta por aproximadamente 5,8 milhões de artigos científicos indexados à plataforma Lattes. Foram identificados pontos de mudança na taxa de crescimento exponencial da produção científica em oito grandes áreas do conhecimento. Tais pontos, estimados através de um modelo de crescimento exponencial, segmentam a produção científica em períodos marcados por taxas constantes de crescimento. A periodização dos resultados facilita a contextualização da produção científica frente às políticas científicas implementadas nas últimas décadas. **Objetivo:** Produzir indicadores para insumos de discussões sobre o desempenho recente da ciência nacional e contribuir com análises que têm como base de dados informações disponibilizadas na Plataforma Lattes. **Metodologia:** Destaca a Plataforma Lattes como banco de dados essencial ao entendimento da ciência brasileira e, a partir da coleta de dados, aplica modelos de

^a Mestranda em Desenvolvimento Econômico na Universidade Estadual de Campinas (UNICAMP). E-mail: dandaraaraujo19@gmail.com.

^b Doutor em Desenvolvimento Econômico pela Universidade Estadual de Campinas (UNICAMP). Docente da Faculdade de Economia da Universidade Federal de Mato Grosso (UFMT). E-mail: roneyfraga@gmail.com.

^c Doutor em Teoria Econômica pela Universidade Estadual de Campinas (UNICAMP). Docente do curso de graduação em Gestão de Agronegócios na Faculdade de Agronomia e Medicina Veterinária da Universidade de Brasília (UnB). E-mail: jaimjunior@gmail.com.

^d Doutor em Política Científica e Tecnológica no Programa de Política Científica e Tecnológica (PPG-PCT) da Universidade Estadual de Campinas (UNICAMP). Pesquisador colaborador do Programa de Pós-Graduação em Política Científica e Tecnológica (PPG-PCT) da Universidade Estadual de Campinas (UNICAMP). E-mail: ludrilvo@gmail.com.

regressão segmentada que compreendem as transformações quantitativas da atividade científica ao longo do período analisado. **Resultados:** Constataram que o total da produção científica brasileira cresceu a uma taxa exponencial anual de 6,61 % entre os anos de 1990 a 2017, porém observando uma desaceleração do crescimento da produção científica nas áreas avaliadas na primeira metade da década de 2000. As áreas de Ciências da Saúde e Biológicas registraram as menores taxas de crescimento. **Conclusões:** Reforçam a relevância da Plataforma Lattes nas avaliações da produção científica nacional, e apontam novos caminhos e indagações para pesquisas futuras que poderão contribuir à compreensão do *status quo* do fazer científico brasileiro.

Descritores: Produção Científica. Regressão Segmentada. Plataforma Lattes. Brasil.

1 INTRODUÇÃO

A pervasividade do conhecimento científico em incontáveis aspectos cotidianos é um fato estilizado capaz de produzir consensos e dissensos entre os mais diversos segmentos da sociedade. Nas palavras de Drori *et al.* (2003) a ciência não pode mais ser confinada a grupos acadêmicos exóticos, já que o conhecimento científico se tornou indispensável para ações de governos e organizações privadas, e está presente em pensamentos e práticas rotineiras de indivíduos comuns. Wilsdon (2011) afirma que tal condição deriva da expansão prévia de todos os níveis de educação, do crescente número de revistas científicas, da especialização contínua da ciência, da maior colaboração internacional entre pesquisadores, dentre outros.

No campo das teorias do crescimento econômico, sabe-se que o acúmulo de capitais tangíveis e o crescimento populacional explicam não mais do que 20% do crescimento observado nos países desenvolvidos (ACEMOGLU *et al.*, 2018; MINNITI; VENTURINI, 2017). Isto porque, os ganhos resultantes de investimentos na produção de conhecimento são maiores para a sociedade do que para os consumidores de bens privados. Como ilustração desse argumento, Nelson (1959) recorda o impacto negativo que o lançamento do satélite russo *Sputinik* provocou na sociedade norte-americana, e como a perda da corrida espacial foi determinante para o incremento dos investimentos em pesquisas científicas nos EUA.

Lane (2010) afirma que medir e avaliar a produção de conhecimento científico se tornaram atividades recorrentes em diversos segmentos da sociedade. Desse modo, estudos bibliométricos passaram a ser largamente

empregados na tomada de decisão sobre a alocação dos investimentos em ciência e tecnologia. No âmbito da pesquisa básica, De Solla Price (1963) pode ser apontado como o pioneiro na investigação de leis empíricas do crescimento da ciência. Segundo o autor, se um segmento suficientemente amplo da ciência for medido de forma razoável, o modo natural de crescimento observado será exponencial, multiplicando-se por um número fixo em intervalos de tempos iguais.

Inspirado no modelo de crescimento da ciência proposto por De Solla Price (1963), o presente estudo estima a taxa de crescimento exponencial de artigos científicos registrados na plataforma Lattes no período 1990 - 2017. Os dados da plataforma Lattes possibilitaram ampliar o contexto da produção científica nacional e produzir resultados que extrapolam os circuitos convencionais de publicações internacionais, sustentados por grandes empresas e editoras. Os objetivos gerais deste artigo são investigar o comportamento de indicadores quantitativos da ciência nacional e avaliar o desempenho recente da produção científica nas grandes áreas do conhecimento.

De modo original, foram identificados pontos de mudança na taxa de crescimento da produção científica nas grandes áreas do conhecimento. Tais pontos, estimados através de um modelo de crescimento exponencial, segmentam a produção científica em períodos marcados por taxas constantes de crescimento. A periodização dos resultados facilita a contextualização da produção científica frente às políticas científicas implementadas nas últimas décadas. Conforme será tratado, sete das oito grandes áreas do conhecimento investigadas foram segmentadas em dois períodos, com o ponto estimado de transição da taxa de crescimento exponencial entre os anos de 2001 e 2004.

A partir dos resultados obtidos nesta pesquisa, pode-se afirmar que a produção científica nacional registrou um período de crescimento acelerado entre 1990 e os primeiros anos da década de 2000, com taxas exponenciais anuais de até 17,6% (Ciências Sociais Aplicadas) e dobrando de tamanho em até cada 4 anos. Após o ponto de mudança, o crescimento da produção passou

a ocorrer de forma menos intensa, sendo necessário até 17 anos para o total de artigos registrados na plataforma Lattes dobrarem de tamanho em algumas áreas do conhecimento.

Além da presente introdução, o artigo está dividido em quatro partes, sendo: i) revisão bibliográfica, que apresenta os aportes da literatura bibliométrica e econômica associadas à produção científica brasileira; ii) procedimentos metodológicos, os quais definem o *modus operandi* do trabalho, destacando a Plataforma Lattes como banco de dados essencial ao entendimento da ciência brasileira, bem como os modelos de regressão segmentada que compreendem as transformações quantitativas desta mesma ciência ao longo do tempo; iii) resultados e discussões, que constata a transformação da produção bibliográfica brasileira de 1990 a 2017, chamando atenção às variações de produtividade dentre as distintas áreas do conhecimento e, finalmente; iv) considerações finais, que reforçam a relevância da Plataforma Lattes e observam características e projeções da produção científica nacional, além de abrir novos caminhos e indagações para pesquisas futuras que poderão também contribuir à compreensão do *status quo* do fazer científico brasileiro.

2 REVISÃO BIBLIOGRÁFICA

Zhang, Powell e Baker (2015) estimaram o crescimento de publicações científicas mundiais no período de 1900 a 2011, com base em artigos indexados no *Science Citation Index Expanded* (SCIE). De acordo com os autores, a capacidade global de gerar conhecimentos científicos atingiu um novo patamar nas últimas três décadas, com as publicações crescendo a taxa anual exponencial de 3,5% entre 1980 - 2011. Além disso, os resultados do trabalho em questão indicam que há em curso um deslocamento do centro de gravidade da produção científica, marcado pelo afastamento da América do Norte, o retorno à Europa e o fortalecimento de países asiáticos.

Bornmann e Mutz (2015) reexaminaram a questão do crescimento da ciência e avaliaram a taxa de crescimento exponencial do número de citações em artigos indexados na base *Web of Science* (WoS). O estudo desenvolveu um

modelo de regressão segmentada específico para identificar períodos da história com taxas similares de crescimento do número de citações. Como resultado, os autores identificaram três fases características do crescimento da ciência global: de menos de 1% até meados do século XVIII, entre 2 e 3% até o fim da segunda guerra mundial, e de 8 a 9% até 2010.

Em uma abordagem regional do crescimento da produção científica, Bornmann, Wagner e Leydesdorff (2014) investigaram o desempenho de Brasil, Índia, China e Rússia (BRIC) no domínio dos artigos mais citados na base WoS no período 1990 – 2010. Os resultados do estudo indicam que, com exceção da Rússia, os países do BRIC aumentaram relativamente ao resto do mundo sua participação entre os artigos mais citados anualmente na base WoS. Em outra pesquisa quantitativa envolvendo países em desenvolvimento, Hayne e Wyse (2018) indicam que houve uma duplicação da taxa de crescimento dos artigos científicos publicados pelos BRICS e indexados na base WoS no período 2008 – 2014.

Cross, Thomson e Sinclair (2017) analisaram o crescimento da pesquisa científica brasileira indexada na base WoS entre os anos de 2011 e 2016. Os autores concluíram que o país ocupa a 13ª posição global de publicações e que os indicadores brasileiros de citações média por artigo e citações dos artigos mais citados melhoraram no período analisado. É importante frisar que avaliações da produção científica baseada em publicações de língua inglesa e em indicadores “*ISI-style*” (*Institute of Scientific Information*) não contabilizam parcela relevante da produção científica nacional. Rodrigues *et al.* (2015) indicam que apenas 314 periódicos científicos brasileiros estavam indexados nas bases Scopus ou WoS em 2015. Para Mugnani *et al.* (2019), é preocupante o fato de que na avaliação de programas de pós-graduação no Brasil as bases de dados bibliográficas comerciais representam quase a totalidade dos critérios de produção científica. Segundo os autores, estudos bibliométricos precisam delinear um contexto mais amplo da produção científica nacional, iluminando espaços de discussão excluídos da chamada corrente principal.

3 PROCEDIMENTOS METODOLÓGICOS

A apresentação dos procedimentos metodológicos empregados no artigo compreende dois subtópicos. O primeiro, aborda às técnicas de raspagem e transformação dos dados sobre a produção científica nacional que constam na Plataforma Lattes. Além disso, o item aborda os critérios adotados para classificar a produção científica em grandes áreas do conhecimento. O segundo subtópico demonstra os modelos de crescimento exponencial e a regressão segmentada utilizados para avaliar o desempenho recente da produção científica nacional em períodos marcados por uma taxa similar de crescimento exponencial.

3.1 EXTRAÇÃO, TRATAMENTO E BASE DE DADOS DA PLATAFORMA LATTES

A plataforma Lattes é reconhecida internacionalmente como exemplo de boas práticas no registro da produção científica (PERLIN *et al.*, 2017). O presente artigo teve como ponto de partida a coleta e o tratamento das informações sobre a produção científica nacional no período 1990 a 2017, disponibilizadas na plataforma Lattes. Os dados foram coletados em junho de 2018 e correspondem a informações da categoria artigos em periódicos. Não foram consideradas as produções de artigos em Anais de Congressos, Livros e Capítulos de Livros. Embora a plataforma Lattes tenha sido lançada em 1998 pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ), os pesquisadores alimentam seus currículos com toda sua vida acadêmica, inclusive com artigos publicados antes de 1998, quando faz-se pertinente.

A plataforma Lattes disponibiliza os dados dos currículos em formato XML - *Extensible Markup Language*. XML é uma linguagem de marcação recomendada pelo W3C - *World Wide Web Consortium* para apresentar dados legíveis para humanos e para máquinas (ABITEBOUL; BUNEMAN; SUCIU, 2000; SPERBERG; McQUEEN, 2000). Após baixar os currículos da plataforma Lattes, os arquivos XML foram importados para o software R (R CORE TEAM, 2020) e transformados em tabelas (*data.frame*), um formato adequado para leitura e tratamento dos dados.

Para baixar os dados da plataforma Lattes foram utilizadas técnicas de Raspagem de Dados (MITCHELL, 2018) auxiliadas pelo software *Captchas Negated by Python reQuests* – CNPQ, conforme Souza (2018). A obtenção dos currículos se torna computacionalmente custosa devido a necessidade de resolver dois *captchas* para baixar cada arquivo. O Lattes dispõe de mais de 5 milhões de currículos cadastrados, que representam aproximadamente 200 GB. Com isso, o processo de resolver um *captcha* para acessar cada currículo, e resolver um segundo *captcha* para exportar os dados dos currículos em formato XML, pode demorar algumas semanas, dependendo das configurações dos computadores utilizados para realizar esta tarefa.

Depois de exportar os arquivos em formato XML, o próximo passo da metodologia foi realizar a leitura e transformação para um formato de fácil manuseio dos dados. O pacote de *R getLattes*, disponibilizado por Souza e Sabino (2020), foi utilizado para importar as informações contidas nos currículos em formato XML e salvá-las em um formato de fácil leitura. A manipulação dos dados demandou conhecimentos de *shell script*, mesmo para realizar tarefas simples como mover os arquivos de um diretório para outro.

No processo de tratamento dos dados é importante compreender que os usuários alimentam seus currículos na plataforma Lattes. Assim, as rotinas para identificar, ordenar e analisar os dados da produção científica são orientadas para o pesquisador. Para eliminar a possibilidade de duplicação na contagem dos dados da produção de artigos em periódicos, que pode ocorrer em trabalhos com mais de um autor, foi necessário pesquisar se mais algum elemento da base de dados tem o mesmo artigo cadastrado com o mesmo título, ano de publicação, e ISSN - *International Standard Serial Number* da revista. O procedimento permite encontrar a quantidade de artigos publicados por ano, contabilizando cada artigo apenas uma vez.

A classificação dos artigos por grande área do conhecimento utilizou informações contidas no campo áreas de atuação do currículo Lattes, registradas pelos pesquisadores. Cada artigo da base de dados foi contabilizado segundo as áreas de atuação declaradas pelo pesquisador na plataforma Lattes, sendo que os pesquisadores podem participar em mais de uma grande área. Com isso,

foi considerada a participação absoluta dos pesquisadores nas grandes áreas do conhecimento. Por exemplo, um pesquisador com atuação em estatística e economia, caso tenha registrado cinco artigos científicos, estes serão contabilizados nas áreas de Ciências Exatas e da Terra, a qual pertence estatística, e Ciências Sociais Aplicadas, a qual pertence economia.

A dissociação das publicações por áreas do conhecimento atendeu a classificação do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), que apresenta nove grandes áreas: 1) Ciências Exatas e da Terra, 2) Ciências Biológicas, 3) Engenharias e Computação, 4) Ciências da Saúde, 5) Ciências Agrárias, 6) Ciências Sociais Aplicadas, 7) Ciências Humanas, 8) Linguística, Letras e Artes, e 9) Outros.

Optou-se pela exclusão da grande área Outros na base de dados. De acordo com a tabela de áreas do conhecimento, disponibilizada pelo CNPq, 23 campos de estudos fazem parte da grande área Outros, que compreendem áreas com bases de conhecimento distintas, como por exemplo, Desenho de Moda, Engenharia Cartográfica e Diplomacia. Não obstante, no processo de tratamento da base de dados utilizada nesse trabalho, arquivos com informações ausentes ou com erros no registro das áreas de atuação foram contabilizados na grande área Outros.

Vale ressaltar que as atividades metodológicas e de análises presentes neste artigo que utilizam material coletado na Plataforma Lattes possuem respaldo legal, garantido pelo CNPq via “políticas de privacidade do currículo Lattes” que observam a possibilidade de terceiros coletarem a informação para criação e análise de indicadores científicos e/ou de política científica e tecnológica (CNPq, 2020).

3.2 MODELO DE CRESCIMENTO EXPONENCIAL E REGRESSÃO SEGMENTADA

A lei do crescimento exponencial da produção científica provém do fato de que, a qualquer momento, a taxa de crescimento é proporcional ao tamanho da população ou à magnitude total atingida. Tague *et al.* (1981) afirmam que o crescimento da produção científica pode ser comparado ao crescimento dos juros compostos, no qual a taxa de crescimento em qualquer momento

corresponde a uma porcentagem fixa do valor atual. Segundo os autores, esse tipo de crescimento é descrito matematicamente pela seguinte função exponencial:

$$F(t) = ae^{b(t-t_0)} \quad (1)$$

No qual $F(t)$ é o tamanho da produção científica no período t ; a é o tamanho inicial da produção; t é o tempo corrente; t_0 é o tempo inicial, e b é a taxa de crescimento. Segundo Egghe e Rao (1992), b pode ser chamado de parâmetro malthusiano, e expressa a proporcionalidade do crescimento ao tamanho da produção científica $F(t)$. No presente artigo, o modelo de crescimento empregado para avaliar a produção científica nacional foi estimado pela seguinte função exponencial:

$$y(t) = b_0 * e^{(b_1 * (t-1990))} \quad (2)$$

Na qual o intercepto b_0 é igual a y_0 , o total de artigos registrados na plataforma Lattes no ano de 1990, e representa o início da série; t representa o ano corrente que vai de 1990 a 2017. Conforme apontam Bornmann e Mutz (2015), o modelo (2) pode ser formulado pela equação diferencial $y'(t) = b_1 * y(t)$ em que b_1 é a constante de crescimento e t é o ano de referência. O presente artigo utilizou o modelo 2 para calcular a taxa de crescimento exponencial da produção nacional agregada e desagregada por grandes áreas do conhecimento. Nesse exercício, a taxa de crescimento em termos percentuais foi obtida pela expressão $(exp(b_1) - 1)$, e o tempo necessário para que o número de publicações dobre de tamanho foi calculado usando $t = \frac{\ln(2)}{b_1}$, expresso em anos.

A regressão segmentada de Bornmann e Mutz (2015) foi aplicada aos dados da plataforma Lattes para estimar os momentos em que a produção científica das grandes áreas passou por mudanças na tendência. A metodologia de segmentação desenvolvida por Bornmann e Mutz (2015, p.6) apoiou-se em estudos bibliométricos anteriores e pode ser descrita pelo seguinte modelo

matemático (BRUSILOVSKIY, 2004; VAN RAAN, 2000; LERMAN, 1980; CARLETON; McGEE, 1970):

$$\begin{aligned} & \text{IF } year \leq a_1 \text{ THEN } \log_y = b_0 + b_1 * year + \varepsilon \\ & \text{ELSE IF } year \leq a_2 \text{ THEN } \log_y = b_0 + b_1 * a_1 + b_2 * (year - a_1) + \varepsilon; \\ & \text{ELSE } \log_y = b_0 + b_1 * a_1 + b_2 * (a_2 - a_1) + b_3 * (year - a_2) \\ & \varepsilon \sim i.i.d. N(0, I\sigma^2). \end{aligned} \tag{3}$$

No modelo (3) a variável dependente é o logaritmo do número de artigos registrado na plataforma Lattes, separado por grande área, em cada ano da série. O parâmetro b_k é a constante de crescimento e a_k o ano de mudança na taxa de crescimento da produção. O cálculo da taxa de crescimento percentual se dá pela expressão $exp(b_1 * t)$, e o tempo necessário para o número de publicações dobrarem por $t = \frac{\ln(2)}{b_1}$.

Os resultados do modelo (3) evidenciam os períodos de aumento ou redução da taxa de crescimento da produção nas grandes áreas. O método adotado para ajustar as curvas de regressão segmentada aos pontos de mudança da taxa de crescimento exponencial seguiu o empregado por Bornmann e Mutz (2015, p. 6-7) e utilizou dois critérios: a) inspeção visual; b) análise estatística. A inspeção visual identificou mudanças na inclinação da curva de crescimento, e com a análise estatística constatou-se que a segmentação adotada foi capaz de explicar 95% da variação anual do total de publicações em cada segmento. Por fim, a análise de regressão segmentada adotou o procedimento *PROC NLIN* do software SAS (SAS, 2011), e a análise da taxa de crescimento foi realizada no software R (R CORE TEAM, 2020).

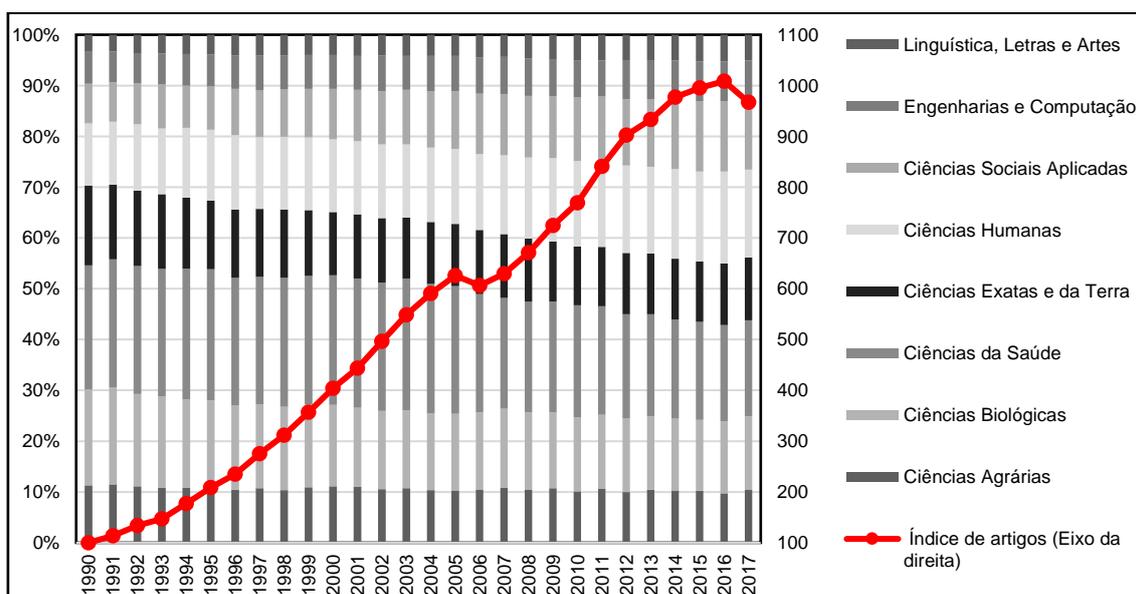
4 RESULTADOS E DISCUSSÃO

O trabalho avaliou 5.647.127 currículos cadastrados na plataforma Lattes até a data de formação da base de dados, que registraram na plataforma o total de 5.886.968 de artigos em periódicos entre 1990 – 2017. Especificamente no período mais recente, 2011 – 2016, foram registrados 2.195.749 artigos científicos nos currículos Lattes dos pesquisadores. A título de comparação, Cross, Thomson e Sinclair (2017) contabilizaram o total acumulado de 250.680

artigos brasileiros indexadas na *Web of Science* no mesmo intervalo de tempo. A grande divergência entre o número de artigos em periódicos registrados na plataforma Lattes e na WoS reforça a necessidade de incentivar a internacionalização da produção científica nacional e lança questionamentos acerca de avaliações do desempenho da pesquisa científica brasileira que têm como base única as publicações indexadas na *WoS*.

A Figura 01 expressa a evolução do total de artigos científicos registrados na plataforma Lattes e a divisão da produção científica entre oito grandes áreas do conhecimento no período 1990- 2017. A linha vermelha captura o crescimento linear da produção científica nacional, e tem como base o número de artigos registrados na plataforma Lattes e publicados no ano de 1990. As barras verticais ilustram a participação de cada grande área do conhecimento no total de artigos em periódicos registrados.

Figura 01 - Artigos registrados na plataforma Lattes no período 1990 – 2017.



Fonte: CNPq – Diretório dos Grupos de pesquisa, elaborado pelos autores (2020).

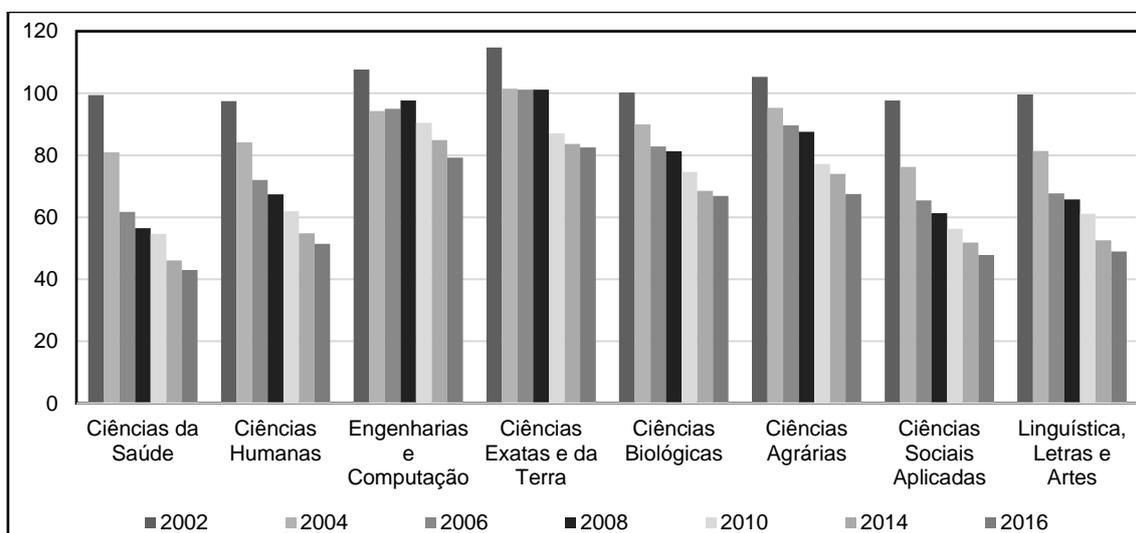
Constatou-se que a produção científica brasileira anual saltou de aproximadamente 40 mil artigos em 1990 para mais de 391 mil em 2016, maior valor da série. Nesse período, o número de artigos publicados por milhão de habitantes no país cresceu aproximadamente 500%, saindo de 340 publicações em 1990 para mais de 1.800 em 2017. Em apenas duas ocasiões, 2006 e 2017,

o total de artigos publicados foi menor do que no ano anterior. No entanto, pode-se constatar que há uma desaceleração da produção nacional a partir de 2014.

Do ponto de vista da produção por grandes áreas do conhecimento, o período 1990 -2017 foi marcado por mudanças no *ranking* das áreas com maior número de publicações anuais. Conforme exposto na Figura 01, Ciências Sociais Aplicadas, Ciências Humanas, Linguística, Letras e Artes fortaleceram suas participações na produção científica nacional, em detrimento das áreas Ciências da Saúde, Ciências Biológicas e Ciências Exatas e da Terra. No mesmo período, as áreas de Engenharias e Computação e Ciências Agrárias mantiveram estáveis suas participações no total de publicações.

A Figura 02 demonstra a evolução no índice de publicações por pesquisador nas grandes áreas do conhecimento, tendo como base o total de artigos registrados na plataforma Lattes e o número de pesquisadores registrados no diretório de grupos do CNPq no ano de 2000. Ao final do período de avaliação, 2016, todas as áreas anotaram redução de pelo menos 20% no número de publicações por pesquisador.

Figura 02 - Evolução do índice de publicações por pesquisador nas grandes áreas.



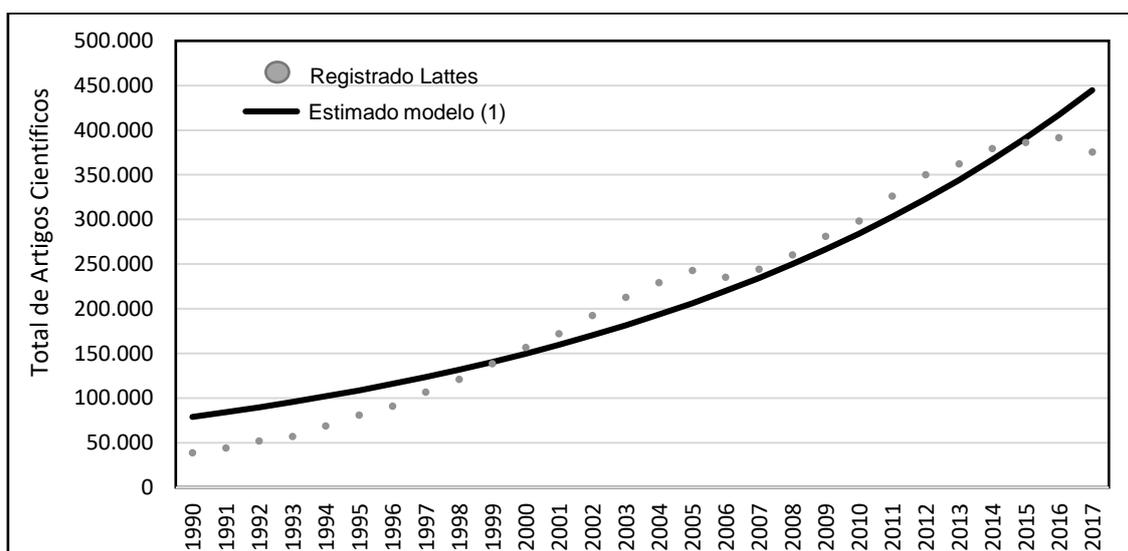
Fonte: CNPq – Diretório dos Grupos de pesquisa, elaborado pelos autores (2019).

Nas áreas de Ciências da Saúde, Ciências Sociais Aplicadas e Linguística, Letras e Artes, o número de publicações por pesquisador caiu pela metade no período em tela. As explicações para a queda acentuada do índice

de artigos por grande área carecem de análises mais aprofundadas, uma vez que diversos fatores podem ter influenciado o resultado negativo. Como por exemplo, a alocação dos investimentos em infraestrutura de pesquisa, a intensidade de parcerias com organizações privadas, a existência de fundos ou programas específicos, os incentivos à qualificação, renovação e reposição do contingente de pesquisadores.

A Figura 03 apresenta os valores da produção científica nacional estimados pelo modelo (2) e os valores extraídos da plataforma Lattes, sendo os valores estimados expressos pela linha contínua. Os resultados indicam que a taxa exponencial de crescimento anual da ciência brasileira entre 1990 e 2017 foi de 6,61 % e o tempo necessário para o número de publicações dobrar de tamanho aproximadamente 11 anos. A curva estimada segue o formato exponencial e, a despeito da desaceleração observada a partir de 2014, não apresenta indícios de saturação ou limites para o seu crescimento. A comparação com os resultados de pesquisas anteriores permite inferir que a produção científica brasileira recente cresceu em ritmo mais intenso que a produção mundial (Zhang, Powell & Baker, 2015).

Figura 03 - Crescimento exponencial da produção científica brasileira 1990-2017



Fonte: Plataforma Lattes, elaborado pelos autores (2019).

Na Tabela 01, os resultados do modelo (2) expressam o crescimento exponencial da produção científica de cada grande área do conhecimento no período 1990 - 2017. Pode-se constatar que a área Ciências Sociais Aplicadas

obteve a maior taxa de crescimento anual: 8,12%. Levando aproximadamente o prazo de oito anos para dobrar de tamanho. As áreas de Ciências Biológicas e Ciências da Saúde registraram as menores taxas de crescimento, respectivamente 5,87 % e 5,19% com o tempo necessário para duplicação das publicações em torno de 12 e 13 anos. As áreas de Ciências Agrárias e Ciências Exatas e da Terra apresentaram taxas de crescimento um pouco menor daquela observada para a produção nacional, 6,26%, 6,19% e 6,61% respectivamente.

Tabela 01 - Taxa de crescimento por área do conhecimento 1990 – 2017.

Área do conhecimento	Taxa de crescimento anual (%)	Tempo para dobrar o número de publicações (Anos)
Ciências Agrárias	6,26	11
Ciências Biológicas	5,87	12
Ciências da Saúde	5,19	13
Ciências Exatas e da Terra	6,19	11
Ciências Humanas	7,66	9
Ciências Sociais Aplicadas	8,12	8
Engenharias e Computação	7,46	9
Linguística, Letras e Artes	7,79	9

Fonte: Plataforma Lattes, elaborado pelos autores (2019).

Segundo Mueller (2005), a árvore do conhecimento disponibilizada pelo CNPq agrupa a ciência de acordo com a especialidade científica e tecnológica que cada área possui, e surge principalmente por conta da diferenciação dos canais que cada campo de pesquisa utiliza para publicar seus resultados. Diversos aspectos específicos às grandes áreas podem ter influenciado o desempenho da produção científica recente, como por exemplo, a intensidade de programas de financiamento, o número de periódicos, a quantidade de cursos de graduação e pós-graduação, o número de docentes, pesquisadores e discentes, as metodologias de pesquisa, dentre outros.

A Tabela 02 expressa os resultados obtidos pelo modelo (3), que segmentou em diferentes períodos o crescimento da produção científica nas grandes áreas do conhecimento. Isto é, a regressão segmentada calculou os períodos em que ocorreram alterações na tendência de crescimento da produção científica em cada grande área do conhecimento.

Tabela 02 - Taxa de crescimento por área do conhecimento 1990 – 2017.

Parâmetro	Estimativa	Desvio padrão	95% Intervalo de confiança	% Taxa de crescimento	Anos para dobrar de tamanho
Ciências Agrárias 1990 - 2017					
a ₁	2001.6	0.3437	2000.9 - 2002.3		
b ₁	0.1355	0.00339	0.1285 - 0.1425	14,5	5,1
b ₂	0.0449	0.00220	0.0404 - 0.0494	4,6	15,4
Ciências Biológicas 1990 - 2017					
a ₁	2002.2	0.3145	2001.6 - 2002.9		
b ₁	0.1191	0.00244	0.1140 - 0.1241	12,6	5,8
b ₂	0.0399	0.00197	0.0359 - 0.0440	4,1	17,4
Ciências da Saúde 1990 - 2017					
a ₁	2003.4	0.4163	2002.5 - 2004.2		
a ₂	2006.1	0.9081	2004.2 - 2008.0		
b ₁	0.1359	0.00337	0.1289 - 0.1429	14,6	5,1
b ₂	(0.0349)	0.0359	(0.1094) – (0.0395)	-3,4	-19,9
b ₃	0.0302	0.00484	0.0201 - 0.0402	3,1	23,0
Ciências Exatas e da Terra 1990 - 2017					
a ₁	2001.7	0.3392	2001.0 - 2002.4		
b ₁	0.1201	0.00274	0.1145 - 0.1258	12,8	5,8
b ₂	0.0457	0.00178	0.0420 - 0.0493	4,7	15,2
Ciências Humanas 1990 - 2017					
a ₁	2001.1	0.4842	2000.1 - 2002.1		
b ₁	0.1543	0.00484	0.1443 - 0.1643	16,7	4,5
b ₂	0.0638	0.00314	0.0574 - 0.0703	6,6	10,9
Ciências Sociais Aplicadas 1990 - 2017					
a ₁	2002.5	0.3696	2001.7 - 2003.2		
b ₁	0.1618	0.00354	0.1545 - 0.1691	17,6	4,3
b ₂	0.0637	0.00285	0.0579 - 0.0696	6,6	10,9
Engenharias 1990 - 2017					
a ₁	2001.4	0.3713	2000.6 - 2002.1		
b ₁	0.1502	0.00370	0.1425 - 0.1578	16,2	4,6
b ₂	0.0593	0.00240	0.0543 - 0.0642	6,1	11,7
Linguística, Letras e Artes 1990 - 2017					
a ₁	2001.1	0.5132	2000.1 - 2002.2		
b ₁	0.1582	0.00525	0.1474 - 0.1690	17,1	4,4
b ₂	0.0655	0.00340	0.0585 - 0.0726	6,8	10,6

Fonte: Plataforma Lattes, elaborado pelos autores (2019).

Na primeira coluna da Tabela 02, o parâmetro a_n informa os momentos estimados pelo modelo (3) em que ocorreram mudanças na taxa de crescimento, e o parâmetro b_n informa a taxa anual de crescimento exponencial registrada em

cada segmento. Com base nesses dados, pode-se inferir que no quarto mês de 2001 ocorreu uma mudança na tendência de crescimento da produção científica na área Engenharias e Computação ($a_1=2001.4$), e a taxa de crescimento anual de 17,1% registrada no intervalo 1990 – 2001.4, caiu para 6,8% ao ano no intervalo 2001.5 - 2017 ($b_1=17,1$; $b_2= 6,8$). Na área de Ciências Agrárias, a mudança no padrão de crescimento da produção científica também ocorreu em meados de 2001 ($a_1 = 2001.6$), marcada pela redução da taxa de crescimento de 14,5% ao ano no período 1990 – 2001.6 para 4,5% no período 2001.7 – 2017. A grande área de Ciências da Saúde foi a única que apresentou duas quebras estruturais ($a_1=2003.4$; $a_2=2006.1$) e, portanto, três fases com taxas anuais de crescimento distintas no período 1990 – 2017 ($b_1=14,6$; $b_2= - 3,4$; $b_3=3,1$). Sendo que entre 2003 e 2006, ocorreu um decréscimo da produção científica na área em questão, resultando em taxa de crescimento negativa.

Os resultados indicam que as oito grandes áreas do conhecimento tiveram padrões de crescimento semelhantes até a data da primeira quebra estrutural, com taxas exponenciais superiores a 10%. No período anterior a quebra estrutural, as taxas anuais variaram entre 12,6% (Ciências Biológicas) e 17,6% (Ciências Sociais Aplicadas), com a duplicação da produção ocorrendo entre quatro e cinco anos.

Destaca-se que em todas áreas do conhecimento as taxas de crescimento no período recente são menores do que as registradas no período anterior a mudança na tendência. Os resultados obtidos vão ao encontro das conclusões destacadas no estudo de Sidone, Haddad e Mena-Chalco (2016) apontam para uma desaceleração da produção nos últimos anos, primordialmente nas grandes áreas de Ciências da Saúde e Ciências Biológicas. Segundo Hayne e Wyse (2018, p. 34), a expansão da produção científica nacional observadas nas décadas de 1990 e 2000 foi impulsionada por aumentos consecutivos nos investimentos em atividades de ciência e tecnologia, que resultaram na abertura e fortalecimento de programas de pós-graduação no país. De forma semelhante, Dias e Moita (2018) afirmam que a expansão dos cursos de doutorado no país coincide com o período de aceleração da produção científica nacional. Kellner (2017) destaca que a redução dos recursos públicos

destinados para os periódicos científicos nacionais observados a partir de 2014, e acentuada em 2016, impactou negativamente o crescimento da produção científica nacional. De modo geral, os resultados do presente artigo reforçaram aqueles encontrados nas pesquisas anteriores sobre o tema.

5 CONSIDERAÇÕES FINAIS

No decorrer desse trabalho foi possível observar a abrangência da plataforma Lattes em relação a ciência brasileira, considerando a quantidade total de currículos cadastrados e a quantidade de publicações científicas encontradas na mesma. Acentuou-se no artigo que as informações da plataforma Lattes fornecem um retrato mais abrangente e fidedigno da produção científica nacional.

No período analisado, a produção científica brasileira cresceu com taxas superiores à média mundial, e as projeções da sua curva exponencial de crescimento não indicaram saturação ou a proximidade de um teto. Porém, a segmentação da produção no tempo, revelou que há em curso uma desaceleração da produção científica nacional, iniciada em meados da década de 2000. Nesse sentido, foi possível constatar que em todas as grandes áreas do conhecimento houve uma redução no número de artigos registrados por pesquisador. O resultado fomentou novas questões relacionadas à evolução da infraestrutura brasileira de pesquisa, que serão investigadas em estudos posteriores. Por exemplo, em áreas do conhecimento mais intensivas em equipamentos de pesquisa e práticas laboratoriais, novos pesquisadores têm encontrado infraestrutura adequada e disponível para desenvolver seus trabalhos?

REFERÊNCIAS

ABITEBOUL, S.; BUNEMAN, P.; SUCIU, D. **Gerenciando dados na Web**. 1.ed. Rio de Janeiro: Campus, 2000.

ACEMOGLU, D.; AKCIGIT, U., ALP, H., BLOOM, N., KERR, W. Innovation, reallocation, and growth. **American Economic Review**, v. 108, n. 11, p. 3450-91, 2018. Doi: <https://doi.org/10.1257/aer.20130470>.

BORNMANN, L.; MUTZ, R. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. **Journal of the Association for Information Science and Technology**, v. 66, n. 11, p. 2215-2222, 2015. Doi: <https://dx.doi.org/10.1002/asi.23329>

BORNMANN, L.; WAGNER, C.; LEYDESDORFF, L. BRICS countries and scientific excellence: A bibliometric analysis of most frequently cited papers. **Journal of the Association for Information Science and Technology**, v. 66, n. 7, p. 1507-1513, 2015. Doi: <https://dx.doi.org/10.1002/asi.23333>

BRUSILOVSKIY, E. The piecewise regression model as a response modeling tool. **NorthEast SAS Users Group (NESUG) Conference Proceedings**. Baltimore: NESUG, p.121 – 134, 2004.

MCZGEE, V.; CARLETON, W.T. Piecewise regression. **Journal of the American Statistical Association**, Issue 331, p. 1109–1124, 1970. Doi: <https://doi.org/10.1080/01621459.1970.10481147>.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO (CNPq). **Termo de Adesão e de Condições de Uso Sistema de Currículos da Plataforma Lattes**. Brasília: CNPq, 2020. Disponível em: https://www.cnpq.br/cvlattesweb/pkg_cv_estr.termo. Acessado em 25 Jan. de 2020.

CROSS, D.; THOMSON, S.; SINCLAIR, A. **Research in Brazil: A Report for CAPES by Clarivate Analytics**. São Paulo: Clarivate Analytics, 2017.

DE SOLLA PRICE, D. J. **O desenvolvimento da ciência: análise histórica, filosófica, sociológica e econômica**. Rio de Janeiro: Livros Técnicos e Científicos, 1976.

DE SOLLA PRICE, D. J. **Little science, big science**. New York: Columbia University, 1963.

DIAS, T. M. R.; MOITA, G. F. Um retrato da produção científica brasileira baseado em dados da Plataforma Lattes. **Brazilian Journal of Information Science: research trends**, v. 12, n. 4, p. 62-74, 2018. Doi: <https://doi.org/10.36311/1981-1640.2018.v12n4.08.p62>.

DRORI, G. S.; MEYER, J.W.; RAMIREZ, O.F.; SCHOFER, E. **Science in the modern world polity: Institutionalization and globalization**. 1.ed. Palo Alto: Stanford University Press, 2003.

EGGHE, L.; RAO, R. Classification of growth models based on growth rates and its applications. **Scientometrics**, v. 25, n. 1, p. 5-46, 1992. Doi: <https://doi.org/10.1007/BF02016845>.

HAYNE, L. A.; WYSE, A. T. S. Econometric Analysis of Brazilian Scientific Production and Comparison with BRICS. **Science, Technology and Society**, v. 23, n. 1, p. 25-46, 2018. Doi: <https://dx.doi.org/10.1177/0971721817744442>.

KELLNER, A. W.A. Editors of Brazilian journals-a hard life that is getting harder! **Anais da Academia Brasileira de Ciências**, v. 89, n. 1, p. 1-2, 2017. Doi: <http://dx.doi.org/10.1590/0001-37652017891>.

LANE, J. Let's make science metrics more scientific. **Nature**, v. 464, n. 7288, p. 488, 2010. Doi: <https://dx.doi.org/10.1038/464488a>.

LERMAN, P. Fitting segmented regression models by grid search. **Journal of the Royal Statistical Society: Series C (Applied Statistics)**, v. 29, n. 1, p. 77–84, 1980. Doi: <https://doi.org/10.2307/2346413>.

MINNITI, A.; VENTURINI, F. The long-run growth effects of R&D policy. **Research Policy**, v. 46, n. 1, p. 316-326, 2017. Doi: <http://dx.doi.org/10.1016/j.respol.2016.11.006>

MITCHELL, R. **Web scraping with Python: Collecting more data from the modern web**. 2.ed. Newton: O'Reilly Media, Inc., 2018.

MUELLER, S.P.M. A publicação da ciência: áreas científicas e seus canais preferenciais. **DataGramZero: Revista de Ciência da Informação**, v.6, n.1, p.1-12, 2005.

MUGNAINI, R.; DAMACENO, R. J. P.; DIGIAMPIETRI, L. A.; MENA-CHALCO, J. P. Panorama da produção científica do Brasil além da indexação: uma análise exploratória da comunicação em periódicos. **Transinformação**, v.31, e190033, p.22-36, 2019. Doi: <http://dx.doi.org/10.1590/2318-0889201931e190033>

NELSON, R. R. The Simple Economics of Basic Scientific Research. **Journal of Political Economy**, vol. 67, no. 3, p.297–306, 1959. Disponível em: <http://jstor.org/stable/1827448>. Acessado em: 22 Jan. 2020.

PERLIN, M. S.; SANTOS, A.P; IMASATO, T.; BORESTEIN, D.; SILVA, S. The Brazilian scientific output published in journals: A study based on a large CV database. **Journal of Informetrics**, v. 11, n. 1, p. 18-31, 2017. Doi: <https://dx.doi.org/10.1016/j.joi.2016.10.008>

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna: R Foundation for Statistical Computing, 2020.

RODRIGUES, R. S.; QUARTIERO, E.; NEUBERT, P. Periódicos científicos brasileiros indexados na Web of Science e Scopus: estrutura editorial e elementos básicos. **Informação & Sociedade**, v. 25, n. 2, p.26 – 42, 2015. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/ies/article/view/117>. Acessado em 23 Jan. 2020.

SAS, S. **STAT 9.3 User's guide**. Cary, New York: SAS Institute Inc, 2011.

SIDONE, O.J.G.; HADDAD, E.A.; MENA-CHALCO, J.P. A ciência nas regiões brasileiras: evolução da produção e das redes de colaboração científica.

Transinformação, v. 28, n. 1, p. 15-32, 2016. <http://dx.doi.org/10.1590/2318-08892016002800002>.

SOUZA, J. F. Captchas Negated by Python reRequests – CNPQ GitHub repositório; GitHub, 2018. Disponível em: <https://github.com/josefson/CNPQ>. Acessado em 24 Jan. 2020.

SOUZA R. F.; SABINO, W. **GetLattes: Read and process data from Lattes curriculum platform**. Zenodo, 2020.

TAGUE, J.; BEHESHTI, J.; REES-POTTER, L. The law of exponential growth: evidence, implications and forecasts. **Library Trends**, v.30, n.1, p.97 – 114, 1981.

SPERBERG-MCQUEEN, C. **Extensible Markup Language (XML) 1.0**. New York: World Wide Web Consortium 1.ed., 2000. Disponível em <http://www.w3.org/TR/2000/REC-xml-20001006>, Acessado em: 22 Jan. 2020.

VAN RAAN, A. On growth, ageing, and fractal differentiation of science. **Scientometrics**, v. 47, n. 2, p. 347–362, 2000. Doi: 10.1023/A:1005647328460.

WILSDON, J. **Knowledge, networks and nations: Global scientific collaboration in the 21st century**. London: Royal Society Policy document, 2011.

ZHANG, L.; POWELL, J.W.; BAKER, D. P. Exponential growth and the shifting global center of gravity of science production, 1900–2011. **Change: The Magazine of Higher Learning**, v. 47, n. 4, p. 46-49, 2015. <https://dx.doi.org/10.1080/00091383.2015.1053777>

EXPONENTIAL PROJECTIONS OF BRAZILIAN SCIENCE: MODELS AND QUANTITATIVE ANALYSIS OF NATIONAL SCIENTIFIC PRODUCTION PUBLISHED IN THE LAST 30 YEARS

ABSTRACT

Introduction: The study examines the growth of Brazilian scientific production in the period 1990 - 2017, based on a sample composed of approximately 5.8 million scientific articles indexed to the Lattes platform. Points of change were identified in the exponential growth rate of scientific production in eight major areas of knowledge. Such points, estimated through an exponential growth model, segment scientific production in periods marked by constant growth rates. The periodization of results facilitates the contextualization of scientific production in the face of scientific policies implemented in recent decades. **Objective:** to produce indicators for inputs for discussions on the recent performance of national science and to contribute to analyzes based on information made available on the Lattes Platform. **Methodology:** highlights the Lattes Platform as an essential database for understanding Brazilian science and, based on data collection, applies segmented regression models that include the

quantitative transformations of scientific activity over the analyzed period. **Results:** They found that the total of Brazilian scientific production grew at an exponential annual rate of 6.61% between the years 1990 to 2017, however observing a slowdown in the growth of scientific production in the areas evaluated in the first half of the 2000s. Health and Biological Sciences areas registered the lowest growth rates. **Conclusions:** reinforce the relevance of the Lattes Platform in assessments of national scientific production and point out new paths and questions for future research that may contribute to the understanding of the status quo of Brazilian scientific makings.

Descriptors: Scientific Production. Segmented Regression. Lattes Platform. Brazil.

PROYECCIONES EXPONENCIALES DE CIENCIA BRASILEÑA: MODELOS Y ANÁLISIS CUANTITATIVO DE PRODUCCIÓN CIENTÍFICA NACIONAL PUBLICADOS EN LOS ÚLTIMOS 30 AÑOS

RESUMEN

Introducción: El estudio examina el crecimiento de la producción científica brasileña en el período 1990-2017, en base a una muestra compuesta por aproximadamente 5.8 millones de artículos científicos indexados a la plataforma Lattes. De manera original, se identificaron puntos de cambio en la tasa de crecimiento de la producción científica en grandes áreas de conocimiento. Dichos puntos, estimados mediante un modelo de crecimiento exponencial, segmentan la producción científica en períodos marcados por tasas de crecimiento constantes. La periodización de resultados facilita la contextualización de la producción científica frente a las políticas científicas implementadas en las últimas décadas. **Objetivo:** producir indicadores que sirvan como insumos para las discusiones sobre el desempeño reciente de la ciencia nacional y contribuir a los análisis basados en la información disponible en la plataforma Lattes. **Metodología:** destaca la Plataforma Lattes como una base de datos esencial para la comprensión de la ciencia brasileña. Basado en la recopilación de este conjunto de datos, aplica modelos de regresión segmentada que incluyen las transformaciones cuantitativas de la actividad científica durante el período analizado. **Resultados:** indican que el total de la producción científica brasileña creció a una tasa anual de 6.61% entre los años 1990 a 2017 y que las áreas de Ciencias de la Salud y Ciencias Biológicas registraron las tasas de crecimiento más bajas en el período. **Conclusiones:** reforzar la relevancia de la Plataforma Lattes y observar las características y proyecciones de la producción científica nacional. Además, abrir nuevos caminos y preguntas para futuras investigaciones que también pueden contribuir a la comprensión del *status quo* de los avances científicos brasileños.

Descriptores: Producción Científica. Regresión Segmentada. Plataforma Lattes. Brasil.

Recebido em: 23.06.2020

Aceito em: 13.12.2020