



Regressão Hierárquica: Um Estudo Sob a Ótica da Bibliometria do Ano 2005 a 2014

Deise Taiana Ávila Dias, Jaime Bettega, Maria Emilia Camargo, Deonir De Toni

RESUMO

Este artigo teve como objetivo investigar a produção científica sobre regressão hierárquica indexada na base de dados *Scopus*, no período de 2005 a 2014. Sendo assim, o estudo assenta numa pesquisa bibliométrica, utilizando-se da estatística descritiva. Evidenciou-se que: o número de artigos publicados no período foi de 11.050; Spertus, J.A. é o autor mais prolífero sobre esta técnica, apresentando 32 documentos; as universidades com maior número de publicações foram a *VA Medical Center*, com 207 publicações e a *University Michigan Ann Arbor*, com 125 publicações; no ano 2014 constatou-se que houve o maior número de artigos publicados, atingindo 1.823 documentos; o país que teve o maior número de publicações foi os Estados Unidos, com 4.790 documentos; por fim constatou-se uma centralidade de publicações na área da Medicina, correspondente a 44,4%. Dentre as contribuições deste trabalho destaca-se, a compreensão do quanto os estudos bibliométricos são ferramentas essenciais para avaliação de periódicos e para desenvolvimento da ciência.

Palavras-chave: Regressão Hierárquica. Produção Científica. Bibliometria.

1 INTRODUÇÃO

Neste estudo serão apresentados os resultados de uma pesquisa que objetivou investigar a produção científica sobre regressão hierárquica indexada na base de dados *Scopus*, no período de 2005 a 2014. As unidades de investigação constituíram-se de artigos que aplicaram a técnica de regressão hierárquica. Sendo assim, para atingir o objetivo deste artigo, optou-se pelo desenvolvimento de uma pesquisa bibliométrica. Como afirma Araújo (2006), as pesquisas bibliométricas têm a finalidade de identificar as características fundamentais das produções científicas de uma área.

Cabe ressaltar que as pesquisas bibliométricas utilizam técnicas estatísticas e matemáticas para apontar particularidades da literatura analisada (ARAÚJO, 2006). Conforme entendimento de Cooper e Lindsay (1998), a análise bibliométrica permite mostrar aos novos pesquisadores os temas já tratados em um determinado campo do conhecimento, também averiguar inconsistências em investigações já realizadas, bem como examinar as lacunas ainda existentes em uma área ou tema específico.

A finalidade da análise de regressão hierárquica é determinar os melhores preditores (variáveis independentes) de um critério (variável dependente) e/ou analisar os efeitos moderadores do relacionamento entre as variáveis (ABBAD; TORRES, 2002). Diante do exposto, ressalta-se a questão central de pesquisa que norteou este estudo: Qual é o perfil da produção científica relacionada à regressão hierárquica, no período do ano 2005 ao ano de 2014?

Espera-se que esta investigação, ao apresentar lacunas ainda não exploradas nos estudos brasileiros e na área da administração, possa contribuir para o desenvolvimento de estudos dessa abordagem no Brasil, bem como possibilite, ao apontar obras essenciais da abordagem e pesquisadores atuantes na área, nortear os acadêmicos que almejam integrar esta área de pesquisa.

Este artigo, no qual se apresenta a pesquisa realizada, encontra-se organizado em cinco seções: (i) a introdução; (ii) o referencial teórico, que trata sobre a bibliometria e a regressão hierárquica, abordando os seus principais pressupostos; o (iii) o método, onde foram descritos



os procedimentos metodológicos adotados; (iv) os resultados encontrados; e por fim, (v) tecem-se as considerações finais do estudo.

2 REFERENCIAL TEÓRICO

2.1 Bibliometria

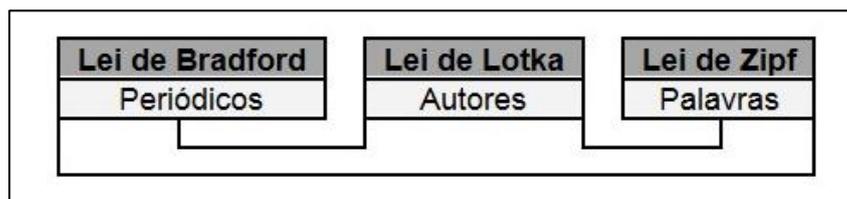
O surgimento da bibliometria deu-se devido à necessidade de estudar e analisar as atividades de produção científica. Pritchard (1969), refere-se a bibliometria como todos os estudos que tentam quantificar os processos de comunicação escrita. Na visão de Guedes e Borschiver (2005), bibliometria é um conjunto de leis e princípios empíricos, os quais corroboram para o estabelecimento dos embasamentos teóricos da ciência da informação. Já Araújo (2006, p. 12) define como uma “técnica quantitativa e estatística de medição dos índices de produção e disseminação do conhecimento científico”.

Araújo (2006) ressalta que, a bibliometria apresenta duas precauções desde a sua origem. O autor cita que umas delas é analisar a produção científica, já a outra é alcançar benefícios práticos e imediatos para as bibliotecas. Um exemplo citado é o desenvolvimento de coleções e a gestão de serviços bibliotecários. Dessa forma, a bibliometria foi:

Inicialmente voltada para a medida de livros (quantidade de edições e exemplares, quantidade de palavras contidas nos livros, espaço ocupado pelos livros nas bibliotecas, estatísticas relativas à indústria do livro), aos poucos foi se voltando para o estudo de outros formatos de produção bibliográfica, tais como artigos de periódicos e outros tipos de documentos, para depois ocupar-se, também, da produtividade de autores e do estudo de citações. (ARAÚJO, 2006, p. 12-13).

Vanti (2002), Bufrem e Prates (2005) e Guedes e Borschiver (2005). evidenciam as três leis bibliométricas mais utilizadas: a Lei de Lotka, a Lei de Zipf e a Lei de Bradford. A Lei de Lotka, também conhecida como Lei do Quadrado Inverso, aborda a medição da produtividade dos autores. A Lei de Zipf ou Lei do Mínimo Esforço refere-se à medição da frequência das palavras em distintos documentos, ocasionando em uma lista ordenada de termos de uma disciplina ou assunto. Por fim, a Lei de Bradford, também conhecida como Lei de Dispersão, ela incide sobre conjuntos de periódicos, instituindo entre eles o núcleo e as áreas de dispersão sobre determinado assunto. Conforme apresentadas na Figura 1.

Figura 1 – Leis Bibliométricas



Fonte: Guedes e Borschiver (2005).

2.2 Regressão Hierárquica

Mediante o entendimento de Jaccard, Turrisi (2003) e Hair et al. (2009), a utilização da regressão é defendida quando se pretende analisar a relação associativa entre uma variável dependente em relação às variáveis independentes, sendo possível avaliar a predição de valores da variável dependente. Sendo assim, são três os tipos principais de análise de regressão múltipla: (i) regressão múltipla padrão; (ii) regressão múltipla hierárquica, ou sequencial; e (iii) regressão estatística (*stepwise* ou *setwise*). Nesta investigação optou-se por tratar de regressão hierárquica.

Abbad e Torres (2002) descrevem a regressão hierárquica como forma de se preparar uma *path analysis*, recomendando que esta regressão é capaz de identificar relações ou



correlações de causalidade entre as variáveis, isto é, os efeitos moderadores. Leia-se que, a regressão hierárquica pode ser utilizada para avaliar os efeitos e a moderação. Conforme equação apresentada na Figura 2.

Figura 2 – Equação Efeito Moderador

$$Y = b_0 + \underbrace{b_1X_1}_a + \underbrace{b_2X_2}_b + \underbrace{b_3X_1X_2}_c$$

Fonte: Hair et al. (2009).

Contudo, cabe ressaltar que como qualquer outra técnica de análise correlacional, a regressão hierárquica não pode identificar causalidade, mas apenas a relação (ou correlação) entre variáveis. Mesmo quando utilizada para testar modelos teóricos que apenas teoricamente indicam uma relação de causa e efeito entre variáveis, a regressão hierárquica está, no máximo, descrevendo padrões de relacionamento entre variáveis (KEPPEL, 1991).

A regressão hierárquica é utilizada em estudos confirmatórios, uma vez que este tipo de análise busca a explicação sobre o relacionamento entre variáveis, descrito em modelos teóricos consistentes. Isto é, em modelos que apresentam um conjunto de proposições empíricas que já indicam a magnitude e direção da relação entre variáveis, mas que, apesar de já terem sido testados, ainda carecem de validação. Neste caso, a ordem de entrada dos preditores na equação de regressão é definida pelo pesquisador, que baseia sua decisão em teorias ou outras pesquisas relacionadas (TORRES, 1999).

No pensamento de Field (2009), na regressão hierárquica, os previsores são selecionados com base em trabalhos anteriores e o pesquisador decide em que ordem eles devem ser colocados no modelo. Como regra geral, previsores conhecidos (de outras análises) devem ser colocados no modelo primeiro pela sua ordem de importância em poder prever a variável de saída. Depois de entrar com todos os previsores conhecidos, o pesquisador pode adicionar os novos ao modelo. Novas variáveis podem ser colocadas de uma única vez, passo a passo ou hierarquicamente (de modo que o novo predictor que se julga mais importante entre primeiro).

2.2.1 Propósitos e Pressupostos da Regressão Hierárquica

O propósito da análise de regressão hierárquica é determinar os melhores preditores (variáveis independentes) de um critério (variável dependente) e/ou examinar efeitos moderadores do relacionamento entre variáveis. Para aplicar a regressão hierárquica, é necessário: (i) os dados devem ser métricos, ou adequadamente transformados; (ii) o pesquisador, antes de estabelecer a equação de regressão, deve decidir qual variável deve ser dependente e quais variáveis serão as independentes. As variáveis individuais devem atender as suposições de: (i) normalidade; (ii) linearidade; (iii) homoscedasticidade e (iv) multicolinearidade.

É necessário estimar a multicolinearidade (correlação) entre as variáveis independentes. A multicolinearidade pode impactar no modelo de regressão, restringindo o poder preditivo de qualquer variável independente que for associada com outras variáveis independentes. Portanto, as variáveis são inseridas na análise em uma determinada sequência. O modelo de regressão é estimado, apresentando como resultado os valores R^2 (ABBAD; TORRES, 2002).

O R^2 é o coeficiente de determinação, o qual indica a proporção de variação total da variável dependente que é explicada conjuntamente pelo modelo de regressão com as variáveis independentes. Cabe destacar que, o coeficiente R^2 pode variar entre 0 e 1. Quanto



maior o valor de R^2 , maior o poder de explicação da equação de regressão. Logo, se o modelo de regressão prever perfeitamente a variável dependente, o valor de R^2 será igual a 1,0 (ABBAD; TORRES, 2002).

A regressão hierárquica examina a relação entre variáveis independentes ou variáveis preditoras (como idade, sexo, severidade da doença) e uma variável dependente (ou variável desfecho) (como óbito, capacidade de exercício). A regressão hierárquica difere da regressão padrão porque um preditor é uma subcategoria de outro preditor. O preditor de nível mais baixo está alojado no preditor de nível mais alto. Por exemplo, em uma regressão que prevê a probabilidade de interrupção de sistemas de apoio à vida em unidades de tratamento intensivo (UTIs) que estão participando de um estudo internacional, a cidade está alojada no país e a UTI está alojada na cidade (GUYATT et al. 2011).

Segundo Barbosa e Fernandes (2000), modelos de regressão são usados para estabelecer a relação entre variáveis explicativas (por exemplo características de alunos, professores, diretores, escolas, etc.) e a variável resposta (desempenho escolar). A estrutura do sistema educacional é organizada hierarquicamente ou em níveis: um grupo de alunos constitui uma turma e um grupo de turmas constitui a escola. Pode-se ainda pensar em níveis superiores de agrupamento, tais como conjuntos de escolas agrupadas por município e estes agrupados em estados. Dados coletados sobre o sistema educacional têm a mesma estrutura. Neste exemplo, identificam-se cinco níveis da hierarquia: alunos, turmas, escolas, municípios e estados.

2.2.2 Estudos Desenvolvidos Com a Regressão Hierárquica

Nos estudos desenvolvidos por Ciptono, Ibrahim e Sulaiman (2010), foi utilizada a regressão hierárquica com mediadores, tendo como finalidade avaliar como a Gestão da Qualidade Total (TQM) – uma filosofia holística de gestão desenvolvida inicialmente por W. Edward Deming, que integra estratégias de melhoria, práticas de gestão e desempenho organizacional – é implementada especificamente em empresas de petróleo e gás que operam na Indonésia. Este estudo foi focado em um conjunto de relações estruturais entre as variáveis para o modelo inteiro.

Sendo assim, para avaliar como as relações estruturais de todo o modelo do estudo (um modelo abrangente de implementação da Gestão da Qualidade Total – TQM) é realmente praticado na indústria de petróleo e gás da Indonésia, dois tipos de questões de pesquisa são abordadas com a análise de mediação usando a técnica de regressão múltipla hierarquizada (CIPTONO; IBRAHIM; SULAIMAN, 2010).

Os resultados encontrados por Ciptono, Ibrahim e Sulaiman (2010), forneceram evidências de que existe uma forte relação positiva entre os fatores críticos de qualidade práticas de gestão e ao desempenho financeiro da empresa mediada pelas três variáveis, ou seja, de classe mundial as práticas da empresa, práticas de excelência operacional, empresa o desempenho não-financeiro.

3 PROCEDIMENTOS METODOLÓGICOS

Ferreira (2011) diz que a bibliometria, é uma técnica de pesquisa que avalia publicações em livros, relatórios e em artigos. Tendo como finalidade quantificar, analisar e avaliar a produção intelectual científica (RAMOS-RODRÍGUEZ; RUÍZ-NAVARRO, 2004), ela é adequada para alcançar o objetivo deste estudo. Sendo assim, justifica-se a utilização da análise bibliométrica por ela fazer um levantamento de artigos, considerando um período prolongado de tempo, também por identificar informações relevantes em periódicos acadêmicos (NERUR; RASHEED; NATARAJAN, 2008), como o agrupamento bibliográfico e a análise das palavras-chave.



Para investigar a produção científica sobre regressão hierárquica, este estudo centrou-se na base de dados *Scopus*, a partir do site www.scopus.com/home.url. A pesquisa foi realizada utilizando-se, como critério de pesquisa, a palavra “Hierarchical Regression”, o período determinado para realização da pesquisa foi do ano 2005 ao ano 2014, o que corresponde a um levantamento longitudinal de 10 anos. A investigação comportou um universo 11.050 artigos.

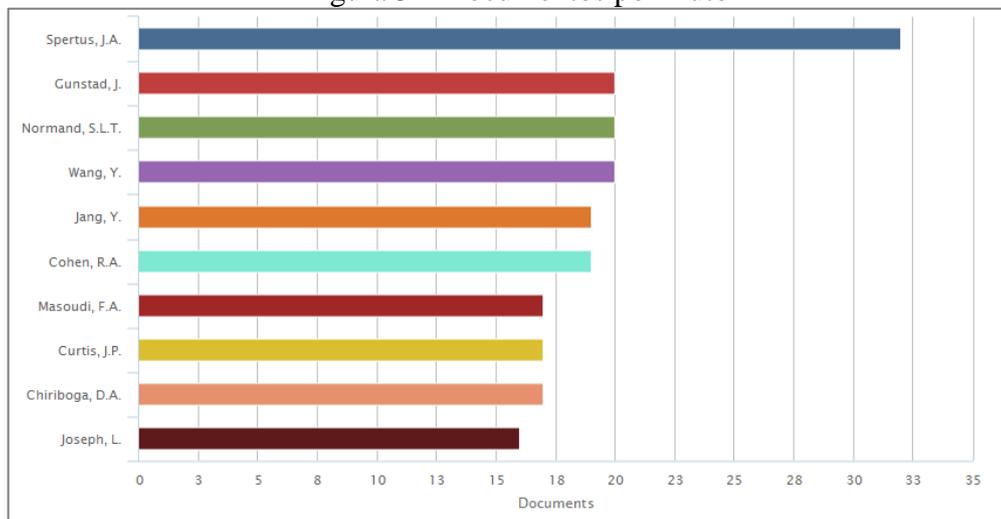
Este estudo fez um levantamento do número de artigos publicados utilizando a regressão hierárquica. Também, os autores que tinham o maior número de publicações. As universidades que apresentaram um maior número de estudos com a utilização da regressão hierárquica. Por conseguinte, qual foi o ano que apresentou menor número de publicações e qual foi o ano que apresentou o maior número de publicações. Por fim, quais foram os países/territórios que apresentaram o maior número de publicações e quais foram as áreas que tiveram o maior número de artigos publicados.

Os resultados a seguir são provenientes da pesquisa descritiva, de natureza quantitativa. No entendimento de Malhotra et al. (2005), a pesquisa descritiva é decorrente de uma análise que está concluindo determinada situação. Seu objetivo principal é descrever algo relacionado ao objeto em estudo, podendo ser um evento ou fenômeno.

4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Com base nos resultados alcançados, pode-se afirmar que os quatro autores que apresentam o maior número de publicações são: Spertus, J.A. com 32 documentos; Gunstad, J. com 20 documentos; Normand, S.L.T. com 20 documentos e Wang, Y. com 20 documentos. Conforme apresentado na Figura 3.

Figura 3 – Documentos por Autor

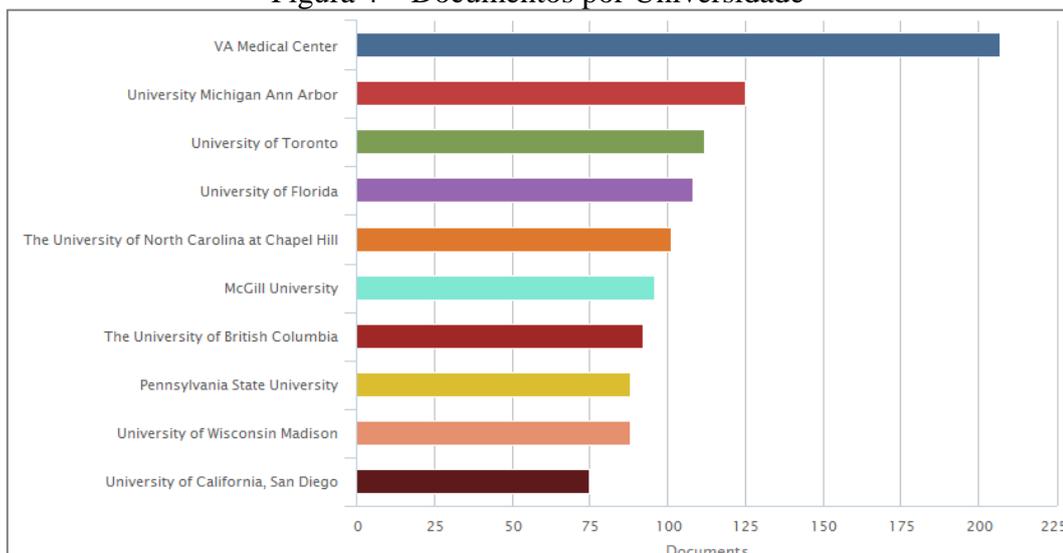


Fonte: Dados provenientes da pesquisa descritiva (2015).

As quatro universidades que apresentaram um maior número de estudos com a utilização da regressão hierárquica foram: VA Medical Center com 207 documentos; University Michigan Ann Arbor com 125 documentos; University of Toronto com 112 documentos e University of Florida com 108 documentos. Conforme apresentado na Figura 4.



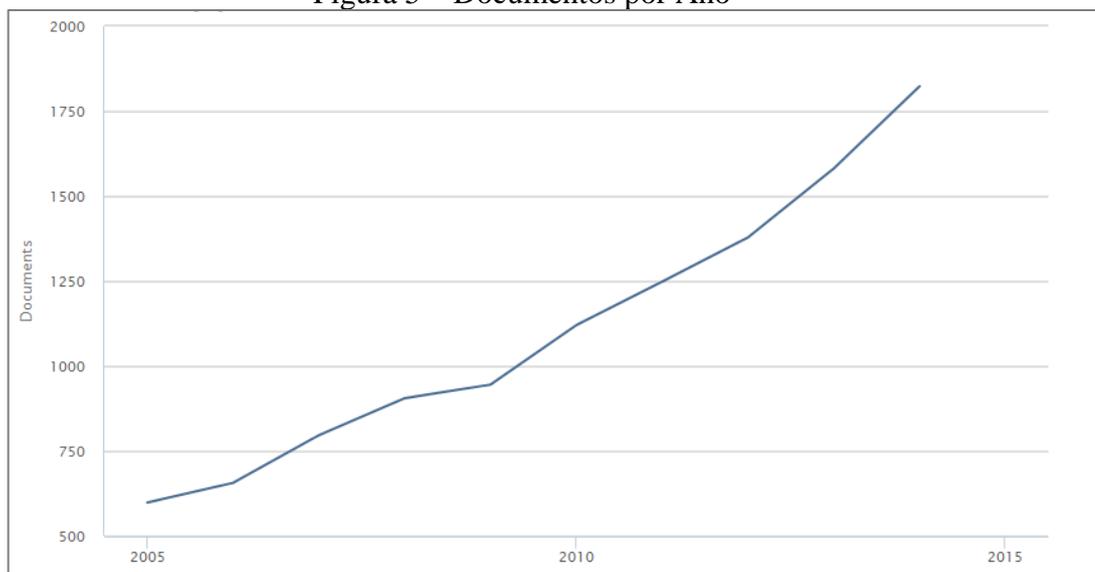
Figura 4 – Documentos por Universidade



Fonte: Dados provenientes da pesquisa descritiva (2015).

Dos resultados alcançados referente a quais os anos que apresentarem o menor e o maior número de documentos, verificou-se que no ano de 2005 foram publicados 598 artigos, no ano de 2014 foram publicados 1.823 artigos. Conforme apresentado na Figura 5.

Figura 5 – Documentos por Ano

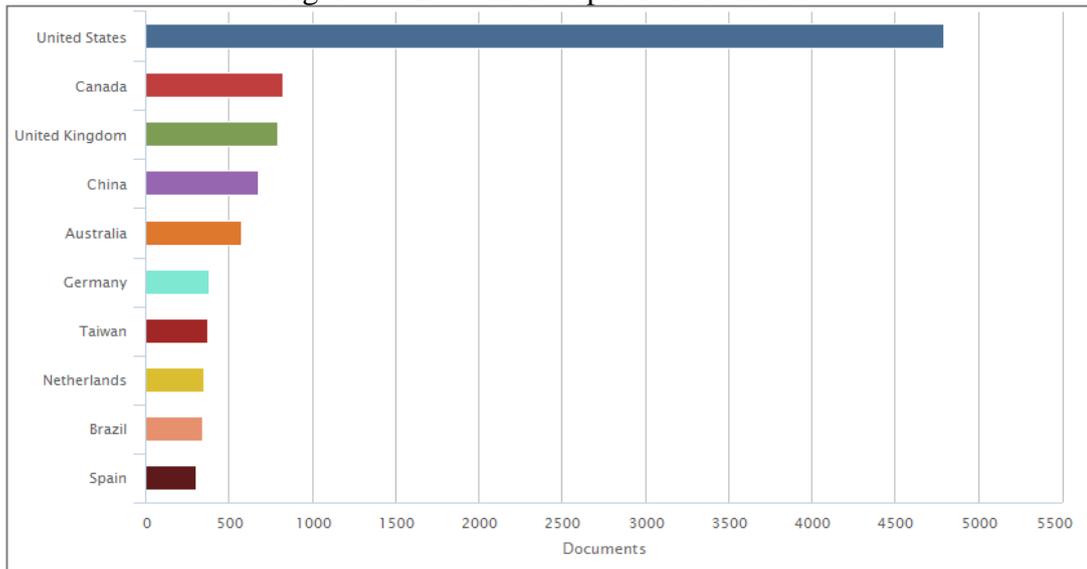


Fonte: Dados provenientes da pesquisa descritiva (2015).

Também, foi feito um levantamento de quais são os quatro países/territórios que apresentaram os maiores números de artigos. Dessa forma, os países/territórios são: Estados Unidos com 4.790 documentos; Canadá com 822 documentos; Reino Unido com 796 documentos e China com 673 documentos. Conforme apresentado na Figura 6.



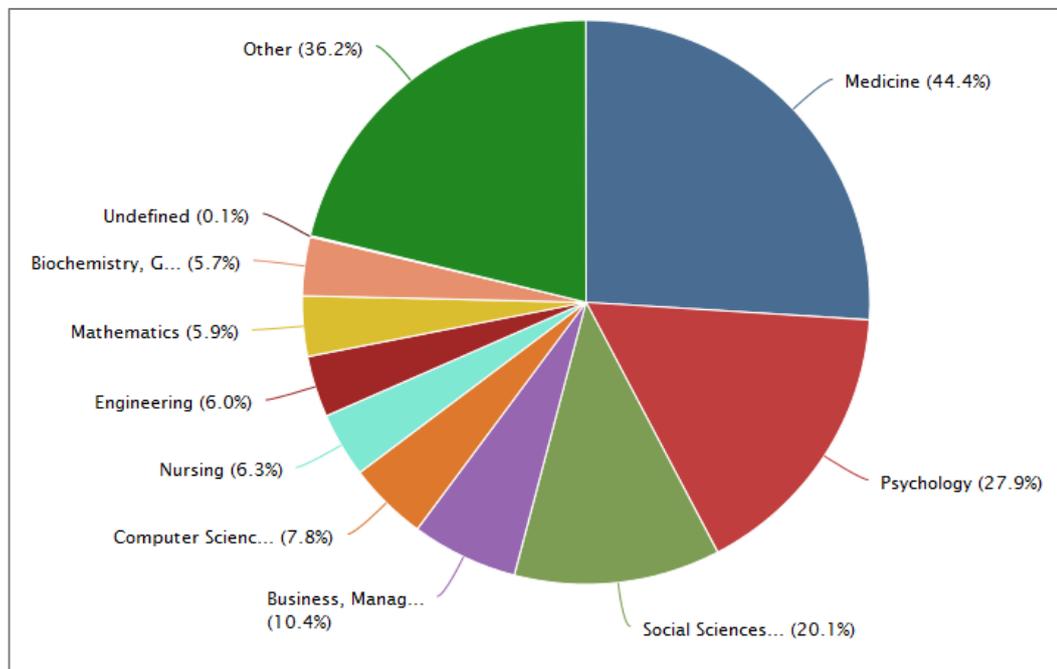
Figura 6 – Documentos por País/Território



Fonte: Dados provenientes da pesquisa descritiva (2015).

Por fim, qual foram as duas áreas que apresentaram o maior número de publicações. Dessa forma, a medicina apresentou 4.908 documentos, resultando em 44,4% do total de publicações. A psicologia apresentou 3.079 documentos, resultando em 27,9% do total de publicações. Conforme apresentado na Figura 7.

Figura 7 – Documentos por Área



Fonte: Dados provenientes da pesquisa descritiva (2015).



5 CONSIDERAÇÕES FINAIS

Este estudo investigou a produção acadêmica científica na base de dados *Scopus*, no período de 2005 a 2014. Para tanto, efetuou-se uma análise bibliométrica em um universo de 11.050 artigos identificados. Dessa forma, observou-se uma predominância de artigos publicados pelo autor Spertus, J.A., com 32 documentos. A universidade com maior número de documentos é VA Medical Center, com 207 documentos.

Nesta investigação também foi identificada que, verificou-se que no ano de 2005 foram publicados 598 artigos, no ano de 2014 foram publicados 1.823 artigos. O que remete ao entendimento que há um crescimento significativo. Logo, é possível afirmar que a regressão hierárquica vem se destacando nas produções científicas, sendo utilizada constantemente.

No que tange ao levantamento de quais foram os países/territórios que apresentaram os maiores números de artigos. Os Estados Unidos lidera o número de publicações, com 4.790 documentos. Já o Brasil, está na 9ª posição, apresentando 340 documentos. Nesse sentido, verifica-se que há interesse dos brasileiros nas publicações com regressão hierárquica, por outro lado, é necessário que os acadêmicos tenham conhecimentos sobre o tema e realizem seus estudos abordando-o.

Por fim, quando investigadas as áreas que apresentaram o maior número de publicações, constatou-se que medicina lidera o número de publicações, com 4.908, isto é, 44,4% do total de publicações. Na sequência, a psicologia apresentou 3.079 documentos, isto é, 27,9% do total de publicações. Com esta análise pode-se afirmar que a área de ciências sociais também está tendo interesse pelas publicações com regressão hierárquica, com 2.217, isto é 20,1%.

De maneira geral, conclui-se que a base de dados *Scopus* reflete o universo da produção científica em distintas áreas. Desse modo, pode auxiliar o acadêmico em fazer um levantamento dos artigos mais relevantes, possibilitando também uma pesquisa exploratória de determinado tema. Assim, ajudando assim a desenvolver, otimizar e socializar determinada área mediante artigos científicos publicados.

Sugere-se como pesquisas futuras que seja feito um levantamento em apenas uma área temática, exemplo ciências sociais. Realizando uma análise do processo de aplicação da regressão hierárquica, vinculado a um determinado número de artigos, para que, com isso, se possa realizar uma análise aprofundada por meio de uma análise de conteúdo.

REFERÊNCIAS

ABBAD, G. S.; TORRES, C. V. **Regressão múltipla stepwise e hierárquica em psicologia organizacional**: Aplicações, problemas e soluções. *Estudos de Psicologia*, Natal, v. 7, p. 19-29, 2002.

ARAÚJO, C. A. **Bibliometria**: evolução história e questões atuais. *Em Questão*, Porto Alegre, v. 12, n. 1, p. 11-32, jan./jun. 2006.

BARBOSA, M. E. F.; FERNANDES, C. **Modelo multinível**: uma aplicação a dados de avaliação educacional. *Estudos em Avaliação Educacional*, São Paulo, n. 22, p. 135-154, jul./dez. 2000.

BUFREM, L.; PRATES, Y. **O saber científico registrado e as práticas de mensuração da informação**. *Ciência da Informação*, Brasília, v. 34, n. 2, pg. 9-25, maio/agosto. 2005.

CIPTONO, W. S.; IBRAHIM, A. Z.; SULAIMAN, A. Mediation analysis using the hierarchical multiple regression technique: a study of the mediating roles of world-class performance in operations. **Gadjah Mada International Journal of Business**, v. 12, n. 2, pp. 139-158, 2010.



Cooper, H. M.; Lindsay, J. J. Research synthesis and meta-analysis. In L. Bickman & D. J. Rog (Eds.), **Handbook of applied social research methods** (pp. 315-342). Thousand Oaks, CA: Sage Publications, 1998.

FIEL, A. **Descobrimo a estatística usando o SPSS**. 2. ed. Porto Alegre : Artmed, 2009.

GUEDES, V.; BORSCHIVER, S. **Bibliometria**: uma ferramenta estatística para a gestão da informação e do conhecimento, em sistemas de informação, de comunicação e de avaliação científica e tecnológica. In: CIFORM – ENCONTRO NACIONAL DE CIÊNCIA DA INFORMAÇÃO, 6., 2005, Salvador. Anais... Salvador: ICI/UFBA, 2005.

GUYATT, G.; RENNIE, D.; MEADE, M.; COOK, D. **Manual para práticas clínica da medicina baseada em evidências**. 2. ed. Porto Alegre: Artmed, 2011.

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.

JACCARD, J.; TURRISI, R. **Interaction effects in multiple regression**. 2nd edition. Thousand Oaks: Sage University Paper, 2003.

KEPPEL, G. **Design and analysis**: A researcher's handbook. Englewood Cliffs: Prentice-Hall, 1991.

NERUR, S. P.; RASHEED, A. A.; NATARAJAN, V. The intellectual structure of the strategic management field: an author co-citation analysis. **Strategic Management Journal**, EUA, v. 29, p. 319-336, 2008.

RAMOS-RODRÍGUEZ, A. R.; RUÍZ-NAVARRO, J. Changes in the intellectual structure of strategic management research: a bibliometric study of the strategic management journal, 1980-2000. **Strategic Management Journal**, EUA, v. 25, p. 981-1004, 2004.

TORRES, C. V. **Leadership style norms among Americans and Brazilians**: Assessing differences using Jackson's return potential model. California School of Professional Psychology, San Diego, 1999.

VANTI, N. A. P. **Da bibliometria à webometria**: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento. *Ciência da Informação*, Brasília, v. 31, n. 2, p. 152-162, maio/ago. 2002.