

# A PRODUTIVIDADE DOS AUTORES NA LITERATURA DE ENFERMAGEM: um modelo de aplicação da lei de Lotka

artigo de revisão

Rubén Urbizagástegui Alvarado\*

## Resumo

Analisa a natureza da distribuição do poder inverso generalizado pelo método dos mínimos quadrados lineares. Descreve passo a passo a forma de aplicação do modelo usando os dados estudados por Estabrooks; Winther & Derksen sobre a literatura de enfermagem. Com  $n = -2.78$ , e  $C = 0.7986$ , ao 0.01 nível de significância, verificou-se que o valor crítico foi de 0.05255, com um desvio máximo de 0.0278, portanto, esta literatura ajusta-se muito bem ao modelo de Lotka.

## Palavras-chave

LEI DE LOTKA  
DISTRIBUIÇÃO DO PODER INVERSO GENERALIZADO  
PRODUTIVIDADE DE AUTORES  
BIBLIOMETRIA  
LITERATURA DE ENFERMAGEM  
INFOMETRIA  
CIENCIOMETRIA

\*Doutorando do Programa de pós-graduação em Ciência da Informação. Escola de Ciência da Informação. Universidade Federal de Minas Gerais. E-mail: ruben@ucr.edu

## I INTRODUÇÃO

A ciência existe também como uma prática social coletiva, e não somente como uma prática individual. Ela é uma realidade social, objetivada em periódicos, instituições, sociedades de especialistas, cátedras universitárias, bibliotecas e orçamentos, especificamente dedicados a apoiar a execução de projetos de pesquisa científica. Um dos problemas que emergem constantemente no mundo acadêmico com relação à prática científica é a produtividade de seus participantes, na forma de publicações, tanto nos seus aspectos quantitativos quanto nos qualitativos. Esta tendência tem dado lugar a reflexões críticas sobre o que agora é considerado como a obrigação de publicar, e a existência de um grupo de significantes contribuintes em qualquer campo do conhecimento. Por conseguinte, poder-se-ia perguntar se a contribuição dos grandes produtores é de maior, igual, ou menor qualidade que a contribuição dos menores produtores. Cattell (1910)

afirmava que não se sabia se o progresso da ciência se devia ao grande número de trabalhadores comuns ou à genialidade de uns poucos. Dennis (1955), analisando a produtividade dos pesquisadores em lingüística, paralisia infantil, gerontologia, geologia e química, observou que 10% dos autores mais prolíficos eram responsáveis por aproximadamente metade de todas as publicações nessas áreas, no entanto, que 60% dos pequenos produtores eram responsáveis por apenas 15% do total das publicações.

O próprio Dennis (1954), estudando o campo da psicologia verificou que os autores mais produtivos eram também aqueles cujos trabalhos eram os mais citados em outros trabalhos acadêmicos do campo da psicologia, logo, haveria uma associação positiva entre a quantidade de trabalhos publicados e o reconhecimento profissional, na forma de "visibilidade" acadêmica. Em consequência, poderia afirmar-se que, entre os autores, há diferenças predeterminadas entre as habilidades e motivações para realizar trabalhos

criativos. Em virtude de diversos fatores sociais, os autores mais produtivos tendem a ser ainda mais produtivos no decorrer do tempo, contudo os autores menos produtivos mostram uma tendência a declinar em produtividade.

Os estudos sobre a produtividade dos autores não são privativos da Biblioteconomia e Ciência da Informação, mas também são realizados por psicólogos e sociólogos, porém em distintas direções. Os psicólogos estão mais interessados em explorar o mundo da criatividade, os fatores cognitivos que fazem possível a existência dos “gênios” e a “inteligência”, enquanto os sociólogos apontam as condições sociais que possibilitam a produção estratificada e desigual na ciência. Os bibliotecários, no entanto, estão mais interessados nas “publicações” (teses, livros, artigos, etc.), como um produto acabado e objetivado, da prática científica.

Sem dúvida, o principal objetivo de qualquer ciência é estabelecer princípios gerais que possam explicar e prever o comportamento dos fenômenos estudados, e por isso a necessidade de se elaborar leis de desenvolvimento e teorias. Em consequência, a prática científica requer a obtenção de dados ou evidências empíricas por um lado, e a elaboração de teorias por outro lado. Por conseguinte, qualquer prática científica deve identificar e delimitar seu objeto de estudo e seus problemas de pesquisa; deve também descobrir leis empíricas que expressem certo tipo de relação entre os fenômenos observados, assim como formular uma estrutura sistemática que contenha as leis empíricas e que estas leis expliquem o fenômeno de maneira racional.

Seguindo os pressupostos epistemológicos estabelecidos anteriormente, os estudos sobre a produtividade dos autores surgiram na década de 1920. Dresden (1922) foi a primeira pessoa a prestar atenção ao fenômeno da produção bibliográfica. Estudou a produção de artigos dos autores ligados à Sociedade Americana de Matemáticas, Seção de Chicago, de 1897 a 1922. Quatro anos mais tarde, Lotka (1926), tentando determinar a parte com que os autores contribuem para o progresso da ciência, contabilizou o número de nomes que apareciam no *Chemical Abstract, 1907-1916*, e no *Auerbach Geschichtstafeln der Physik*, até 1900. Em seguida, traçou a frequência de pessoas que efetuavam 1, 2, 3, etc. contribuições face ao número de 1, 2, 3, etc. contribuintes, com ambas as variáveis em escala logarítmica. Observou que os pontos estavam estreitamente dispersos sobre uma linha reta, tendo uma inclinação de aproximadamente igual a 2,

concluindo, conseqüentemente, que a fórmula geral para essa relação entre a frequência e o número de pessoas que efetuavam  $x$  contribuições era  $x^n y = \text{const}$ , e a proporção dos autores que contribuíam com um único item era de mais ou menos 60 por cento. Inicialmente, o trabalho de Lotka não chamou nenhuma atenção e não foi senão até 1941 (quinze anos depois de sua publicação) que teve sua primeira citação (POTTER, 1981). Seis anos depois, Carr (1932) estudou uma amostra aleatória de 1.000 nomes do índice alfabético do registro de patentes americanas de 1916, e continuou conferindo o registro de cada indivíduo na amostra por dez anos até 1925 e ainda outra vez em 1930. Analisando essa amostra, foi encontrada que consistia de 904 residentes dos Estados Unidos e de 96 residentes de países estrangeiros e possessões insulares, principalmente a Inglaterra, Alemanha, França, Suécia e Suíça. Em dez anos os 1.000 inventores produziram 3,457 patentes com uma média de 1.144 patentes por indivíduo. Somente 13% dos 904 inventores americanos produziram mais de 5 patentes; porém, deste grupo, quase a sétima parte dos inventores produziram 57,5% das patentes, no entanto que 53,2% deles produziram somente uma patente. A produção das patentes parecia seguir a distribuição que Lotka já tinha observado na publicação de trabalhos científicos. Já em 1938, Dufrenoy (1938) examinou o comportamento das publicações dos biólogos e sugeriu que esses dados refletiam a lei de Lotka somente para os pequenos valores das  $n$  contribuições. Hersh (1942), usando uma exaustiva bibliografia publicada em 1939 por H. J. Muller sobre genética de *Drosophila*, analisou o número de artigos publicados em cada ano, desde 1910 até 1938, segundo o número de autores. Quando o número de artigos publicados foram traçados numa escala semilogarítmica diante do tempo decorrido, verificou que os pontos se agrupavam dentro de uma estreita banda com bordas retas e paralelas, ou seja, os dados se ajustavam à simples relação exponencial chamada de lei do interesse composto. Porém, quando o logaritmo do número de autores foram traçados face ao logaritmo do número de artigos creditados aos autores, os dez primeiros pontos (que agrupavam aproximadamente 90% dos autores) se alinhavam em linha reta, com uma inclinação negativa que assemelhava-se à lei de Lotka. Em 1944, Williams (1944), discutindo os dados de Dufrenoy (1938) juntou outros dois dados sobre a produção dos biólogos e, ao analisar seus resultados, constatou que estes eram similares aos de Dufrenoy, com -0.40

e -0.31, como cálculos das inclinações, respectivamente.

Leavens (1953) avaliou os trabalhos dos econométricos usando o *Guide to Econometrica*, de dezembro de 1952. Este guia continha informações pertinentes aos artigos publicados nos primeiros vinte volumes de *Econométrica* (1933-1952). Seus dados produziram resultados similares à curva de Pareto sobre ingressos familiares, os quais têm uma inclinação reta. Esta informação relacionava-se à generalização de Zipf, e pode ser encontrada em diferentes fenômenos econômicos e sociais. O autor discutiu o fato de que o gráfico não se ajustava a uma linha reta em toda a sua extensão. Não obstante, traçando a linha à mão, quase todos os pontos, com exceção dos três últimos, se ajustavam estreitamente à teoria de Pareto (com uma inclinação de -1.5 aproximadamente). Ainda que Leavens tivesse se dado conta de que os artigos não produziam o que deles se esperava, explicou que *Econométrica* era uma entre muitas outras fontes de publicação disponíveis para os 721 autores contribuintes.

O modelo de Lotka tem sido testado em muitas áreas que incluem bases de dados de patentes (OPPENHEIM, 1986), óleos lubrificantes (LÓPEZ CALAFI; SALVADOR; GUARDIA, 1985), educação superior (BUDD, 1988), música popular (COOK, 1989), finanças (CHUNG; COX, 1990), economia (COX; CHUNG, 1991), contabilidade (CHUNG; PAK; COX, 1992), indústria musical (COX; FELTON; CHUNG, 1995), psiquiatria (LÓPEZ-MUÑOZ; RUBIO VALLADOLID, 1995), glândula pineal e melatonina (LÓPEZ-MUÑOZ et al., 1996), biblioteconomia e ciência da informação espanhola (JIMÉNEZ CONTRERAS; MOYA ANEGÓN, 1997), genética (GUPTA; KUMAR; ROUSSEAU, 1998), bibliometria (ALVARADO, 1999), geologia (ALVARADO; CORTÉS, 2002). Porém, os dados destas pesquisas variam muito em extensão e amplitude, indo desde dados tomados de bibliografias exaustivas, como as de entomologia (GUPTA, 1987), pesquisa sobre a batata (GUPTA et al., 1996), até dados compilados de um grupo de periódicos (NATH; JACKSON, 1991; ESTABROOKS; WINTHER; DERKSEN, 2004), e inclusive de um único periódico (CLINE, 1982; ALBEROLA et al., 1985; SÁNCHEZ HERNÁNDEZ et al., 1995; AL-GHAMDI et al., 1998; IVANOV; LUKOVSKAYA, 2003). Alguns estudos usaram dados sobre a história de um único tema (WAGNER-DOBLER; BERG, 1996), mas outros consideraram somente dados sobre poucos anos

(OLIVEIRA; CALDEIRA, 1976; FOX; LASKER, 1983). Deve-se recalcar também que a lei de Lotka tem sido testada em face de muitas compilações de dados e o ajuste nem sempre tem sido bom. Embora este modelo tenha um valor heurístico considerável, agora está claro que o modelo do quadrado inverso como proposto por Lotka (1926) não é mais aplicável a um grande corpo de dados empíricos.

Este ponto não tem sido o suficientemente recalçado na bibliografia publicada, ainda que as revisões do estado-da-arte de Vlachy (1980) e Potters (1981) tenham ajudado a reorientar o interesse dos pesquisadores para avaliar se outros tipos de distribuições podem proporcionar um melhor ajuste dos dados. Isto levou a introduzir-se nas pesquisas modelos estatísticos usados nas ciências naturais, tais como a distribuição hiperbólica, a distribuição logarítmica normal, a distribuição de Yule, a distribuição binomial, a distribuição negativa binomial, a série geométrica, a série logarítmica, a distribuição de Weinbull, a distribuição de Poisson, a distribuição truncada de Poisson e, finalmente, a distribuição de Gauss-Poisson inversa generalizada (URBIZAGÁSTEGUI ALVARADO; OLIVEIRA, 2001; URBIZAGÁSTEGUI ALVARADO, 2002). É importante observar que esses modelos estatísticos não têm sido suficientemente explorados na bibliometria latino-americana, apesar de estarem disponíveis na literatura publicada e serem acessíveis por meio dos textos de ensino. Talvez isto resulte da falta de familiaridade com os modelos estatísticos ou do desconhecimento dos instrumentos matemáticos para testar os dados. Sabe-se que desentranhar complexos modelos estatísticos não é o passatempo favorito de ninguém e, por esta razão, o objetivo deste trabalho é proporcionar um guia didático para analisar e testar a lei de Lotka sobre a produtividade científica dos autores, usando o modelo do poder inverso generalizado pelo método dos mínimos quadrados. Consideramos que este guia é necessário para apoiar a compreensão, adoção, aplicação e difusão deste modelo.

## **2 NATUREZA DA DISTRIBUIÇÃO DO PODER INVERSO GENERALIZADO**

A lei de Lotka é uma distribuição de probabilidades discretas que descreve a produtividade dos autores. Originalmente proposta por Lotka (1926) como um modelo do quadrado

inverso, agora é conhecida como a lei de Lotka, uma forma mais geral chamada de poder inverso generalizado, e que é expressada na forma de:

$$y_x = C x^{-n}, \quad x = 1, 2, \dots, x_{max} \quad (1)$$

onde,  
 $y_x$  é a probabilidade de que um autor faça  $x$  contribuições sobre um assunto  
 $C$  e  $n$  são os dois parâmetros que devem ser estimados dos dados observados.

Os principais elementos envolvidos no "ajuste" do modelo do poder inverso generalizado são os seguintes:

### 3 MEDIÇÃO E TABULAÇÃO

O número de autores  $y_x$  contribuindo com  $x$  artigos em um determinado assunto devem ser organizados em uma tabela de freqüências decrescentes de  $N$  pares  $x$  e  $y$ . As medidas da produtividade dos autores devem levar em consideração todos os autores, incluindo os colaboradores. Aparentemente, Lotka contou somente os primeiros autores porque múltiplos autores eram menos comuns naquele tempo e provavelmente porque essa forma de contagem era mais fácil. Num mundo pós-moderno e de capitalismo avançado, a pesquisa se caracteriza por uma extensiva colaboração, que se reflete nos múltiplos autores de um único trabalho. Se estamos interessados na distribuição da produtividade desses autores, as medidas que não levem em conta essas colaborações, e que não sejam sensíveis a este fenômeno, são inválidas.

#### 3.1 Modelo adotado

O modelo adotado é o poder inverso generalizado, na forma de:

$$y_x = C \left( \frac{1}{x^n} \right) \quad (2)$$

que é a forma que adota a equação (1) quando se elimina desta o signo negativo do expoente  $n$ .

### 4 ESTIMAÇÃO DO PARÂMETRO $N$

O Parâmetro  $n$  é calculado usando o método do mínimo quadrado linear e expressado como

$$n = \frac{N \sum XY - \sum X \sum Y}{N \sum X^2 - (\sum X)^2}$$

onde,  
 $N$  = número de pares de dados observados  
 $X$  =  $\log x$  (base 10)  
 $Y$  =  $\log y$  (base 10)  
 2.4 Estimação do parâmetro  $C$

Para estimar  $C$  se usa a função inversa  $Z$  de Riemann. Para esta estimação, Pao (1986) proporciona uma fórmula de aproximação exata, expressada como:

$$C = \frac{1}{\sum_{x=1}^{P-1} \frac{1}{x^n} + \frac{1}{(n-1) P^{n-1}} + \frac{1}{2 P^n} + \frac{n}{24 (P-1)^{n+1}}}$$

onde,  
 $x$  = é o número de 1, 2, 3, ...  $n$  contribuições por autor  
 $n$  = é o valor do parâmetro estimado pelo método do mínimo quadrado linear  
 $P$  = é o número de pares de dados  $xy$  observados.

#### 4.1 Teste

O teste de ajuste usado é o Kolmogorov-Smirnov (K-S), que é aplicado ao conjunto de valores observados e esperados a um 0.01 nível de significância. Este teste compara a distribuição das freqüências observadas com a distribuição das freqüências calculadas ou teóricas, usando as funções acumuladas de ambas as distribuições.

### 5 APLICAÇÃO DO MODELO DO PODER INVERSO GENERALIZADO PELO METODO DOS MINIMOS QUADRADOS

Para realizar um teste de ajuste da lei de Lotka mediante o modelo do poder inverso generalizado, usando o método do mínimo quadrado linear e aplicando a uma distribuição da produtividade de

autores, recomenda-se acompanhar os seguintes passos. Na exemplificação desses passos vamos utilizar os dados procedentes de um estudo sobre a literatura de enfermagem, realizado por Estabrooks, Winther e Derksen (2004).

### 5.1 Coleta dos dados e distribuição das freqüências observadas

Tome-se uma bibliografia existente (ou elabore uma), cuja cobertura seja extensa (quanto mais extensa em tempo, melhor). Sugere-se uma cobertura de dez anos ou mais. Efetua-se a contagem do número de contribuições feitas por cada autor, incluindo os co-autores. Ordene esses dados numa tabela de oito colunas. Nas duas primeiras colunas indique os valores de  $x$  (o número das contribuições, ex. 1, 2, 3, 4 ... etc.) em ordem crescente, e os valores de  $y$  (o número de autores que tenham feito  $x$  contribuições), ordenados de forma decrescente; portanto, o primeiro  $x$  (uma contribuição) deverá conter o maior número de autores com uma contribuição.

Aqui, a primeira previsão a ser tomada deve ser a de completar as contribuições com os valores das freqüências não observadas. Na opinião de Loughner (1992), trata-se de evitar que existam células vazias que, mais adiante, possam afetar o teste Kolgomorov-Smirnov. Por exemplo, nos dados de Estabrooks, Winther e Derksen (2004) não se observaram autores que produziram 10 colaborações, por isso, essa décima colaboração foi completada com zero autores produtores.

Na terceira coluna indique os valores de  $xy$  (multiplicação de  $x$  por  $y$ ). Esta coluna indica o número total de artigos produzidos pelos autores. Na quarta coluna indique  $\Sigma xy$  (a soma acumulada da multiplicação de  $x$  por  $y$ ). Na quinta coluna indique a porcentagem de  $y$ , ou seja, a porcentagem dos autores. Na sexta coluna indique a  $\Sigma \%y$  (o valor acumulado da porcentagem de  $y$  da quinta coluna). Na sétima coluna a porcentagem do total dos artigos (% de  $xy$  da terceira coluna). Na oitava coluna indique os valores acumulados da sétima coluna. A Tabela 1 mostra essa forma de organização dos dados.

**Tabela 1:** Freqüência observada de contribuições por autor

1	2	3	4	5	6	7	8
No. de contribuições por autor $x$	No. de autores $y$	Total de artigos $x.y$	$Sx.y$	% de autores $\% y$	$S \% y$	% de artigos $\% xy$	$S \% xy$
1	795	795	795	82.6	82.6	62.4	62.4
2	103	206	1001	10.7	93.3	16.2	78.6
3	33	99	1100	3.4	96.7	7.8	86.4
4	12	48	1148	1.2	97.9	3.8	90.2
5	6	30	1178	0.6	98.5	2.4	92.6
6	5	30	1208	0.5	99.0	2.4	95.0
7	2	14	1222	0.2	99.2	1.1	96.1
8	3	24	1246	0.3	99.5	1.9	98.0
9	2	18	1264	0.2	99.7	1.4	99.4
10	0	0	1264	0.0	99.7	0.0	99.4
11	1	11	1275	0.1	99.8	0.9	100.0
Total	962	1275		100.0		100.0	

Esta primeira tabela serve somente para uma observação geral dos valores obtidos. Calcule com ela a produtividade media por autor, descreva a porcentagem de autores com uma

única contribuição, etc., isto é, algum aspecto interessante a ressaltar na distribuição. Por exemplo, nos dados de Estabrooks, Winther e Derksen (2004) existe uma concentração de 83%

dos autores contribuindo com um só artigo. Esta porcentagem é 23% mais alta que os 60% prognosticados pelo modelo do quadrado inverso de Lotka. No lado oposto, temos apenas 2% dos autores que produziram 5 ou mais artigos. A média da produtividade dos autores é de 1.3, com uma variância de 1.02 artigos por autor.

## 6 TABELA DOS MÍNIMOS QUADRADOS

Elabore outra tabela contendo de novo seis colunas. As colunas 1 e 2 contêm os mesmos valores

de  $x$  e  $y$ , conforme descrito no ponto 1. A coluna 3 contém o logaritmo (base 10) dos valores de  $x$  da coluna 1. A coluna quatro contém os valores do logaritmo (base 10) do número de autores  $y$  da coluna 2. A coluna cinco contém os valores da multiplicação dos logaritmos de  $x$  por  $y$  (coluna 3 por coluna 4). A coluna seis contém os valores do logaritmo de  $x$  elevados ao quadrado (coluna 3 elevados ao quadrado). Depois se realiza a somatória dos valores das colunas 2 ao 6. Incluiu-se aqui um exemplo com os dados da literatura de enfermagem, elaborado por Estabrooks, Winther e Derksen (2004). Estas especificações são mostradas na Tabela 2 a seguir:

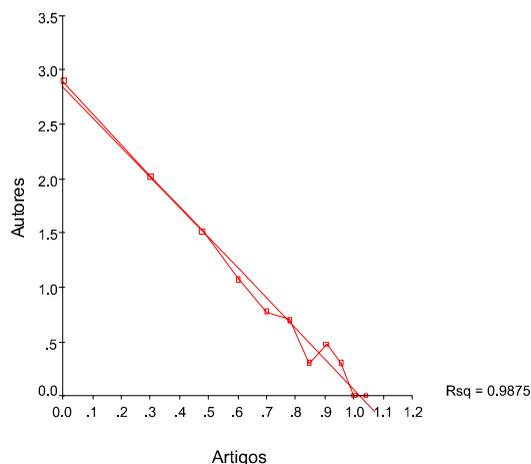
**Tabela 2:** Distribuição dos mínimos quadrados dos dados observados

1	2	3	4	5	6
No de contribuições por autor $x$	No. de autores $y$	$\log x.$	$\log y$	$\log x (\log y)$	$(\log x)^2$
1	795	0.00000	2.90037	.00000	.00000
2	103	0.30103	2.01284	.60592	.09062
3	33	0.47712	1.51851	.72452	.22764
4	12	0.60206	1.07918	.64973	.36248
5	6	0.69897	0.77815	.54390	.48856
6	5	0.77815	0.69897	.54390	.60552
7	2	0.84510	0.30103	.25440	.71419
8	3	0.90309	0.47712	.43088	.81557
9	2	0.95424	0.30103	.28726	.91058
10	0	1.00000	0.00000	.00000	1.00000
11	1	1.04139	0.00000	.00000	1.08450
Total	962	7.60116	10.06720	4.04052	6.29966

### 6.1 Traçado do gráfico

Trace o logaritmo de  $y$  diante do logaritmo de  $x$ . Alguns autores recomendam inspecionar visualmente o traçado para determinar o ponto de corte da linha reta. Este é o chamado método do "Olhómetro". Recomendam também cortar os valores dos grandes produtores que se desviam

demasiado da linha reta de regressão. Outros autores recomendam não realizar nenhum corte e trabalhar com todos os autores e co-autores incluídos na distribuição das frequências. Como podemos ver na Figura 1, com os dados completos (sem corte algum), a regressão linear entre os artigos produzidos e os autores produtores de artigos é de 98.75%



**Figura 1:** Traçado do coeficiente de determinação de Pearson

Agora, se cortamos os dois últimos autores de maior produção, a correlação entre autores e artigos produzidos pode aumentar, mas neste caso a correlação permanece estável (98.44%). Precisamente, Pao (1986, p. 29) recomenda calcular a regressão linear ou coeficiente de determinação  $r^2$  removendo, cada vez, um par de dados dos autores mais produtivos, até atingir a maior porcentagem de  $r^2$ . Este coeficiente de determinação permite estabelecer a quantidade de variação na variável  $y$ , que é explicada pela equação de regressão produzida pela variável  $x$ . O objetivo é encontrar a “melhor” linha de regressão para os dados em estudo. Não obstante, outros pesquisadores recomendam trabalhar com todos os autores, ou seja, sem corte algum. Especialmente porque o cálculo do coeficiente de determinação  $r^2$  pode ser feito facilmente com pacotes estatísticos automatizados, tais como Excel, Notebook, Minitab, SPSS, SAS, Mathematica, Matlab, e outros disponíveis no mercado.

### 6.2 Cálculo do parâmetro $n$

Utilizando os valores da tabela dos mínimos quadrados, elaborados no ponto 2, fazer o cálculo do “melhor” valor da queda de  $n$ . Esta queda é o valor do expoente  $n$  da Lei de Lotka. Para calcular o valor de  $n$ , use a seguinte equação:

$$n = \frac{N \sum XY - \sum X \sum Y}{N \sum X^2 - (\sum X)^2}$$

onde,

$N$  = número de pares de dados  $xy$  observados

$X$  =  $\log x$  (Base 10)

$Y$  =  $\log y$  (Base 10)

Para ilustrar o cálculo do parâmetro  $n$ , vamos usar os dados de Estabrooks, Winther e Derksen (2004), tomando os valores já calculados na Tabela 2. Neste caso, usaremos todos os dados sem corte algum, que é o mais recomendável.

$$n = \frac{11(4.04052) - (7.60116)(10.06720)}{11(6.29966) - (7.60116)^2}$$

$$n = \frac{44.446 - 76.522}{69.296 - 57.778}$$

$$n = \frac{-32.076}{11.518}$$

arredondando,

$$n = -2.78$$

### 6.3 Cálculo do parâmetro $c$

O parâmetro  $C$  representa a porcentagem teórica dos autores que colaboraram com um único artigo ou trabalho na distribuição da produtividade dos autores. Fazer o cálculo de  $C$ , substituindo o valor de  $n$ , obtido na etapa 4, e usando  $P = 11$  (porque são onze pares de dados), na seguinte equação:

$$C = \frac{1}{\sum_{x=1}^{P-1} \frac{1}{x^n} + \frac{1}{(n-1)P^{n-1}} + \frac{1}{2P^n} + \frac{n}{24(P-1)^{n+1}}}$$

onde,

$x$  = é o número de 1, 2, 3, ...  $n$  contribuições por autor

$n$  = é o valor do parâmetro  $b$  estimado no ponto 4.

Neste caso,  $n$  é igual a  $-2.78$

$P$  = é o número de pares de dados observados. Neste caso são onze pares.

O valor de  $P$  variará, dependendo do número de pares  $x$  e  $y$  observados. Se os pares de dados

observados forem 10, 15, 16, etc., o valor de P também deverá ser igual a 10, 15 ou 16. Por exemplo, continuando com os dados de Estabrooks, Winther e Derksen (2004).

$$C = \frac{1}{\sum_{x=1}^{P-1} \frac{1}{x^{2.78}} + \frac{1}{(2.78-1)(11)^{2.78-1}} + \frac{1}{2(11)^{2.78}} + \frac{2.78}{24(11-1)^{2.78+1}}}$$

O denominador desta equação está formado pela soma de 4 equações diferentes, assim, vamos calculá-las separadamente, uma por uma.

a) Primeira equação

A equação  $\sum_{x=1}^{P-1} \frac{1}{x^{2.78}}$ , que é a primeira parte

do denominador, indica um somatório, que vá desde  $x = 1$  até  $P - 1$ . Se  $P = 11$ , então,  $11 - 1 = 10$ . Portanto, a somatória vá de  $x = 1$  até  $P = 10$ . Esta equação é obtida da maneira seguinte:

$$\sum_{x=1}^{P-1} \frac{1}{x^{2.78}} = \frac{1}{1^{2.78}} + \frac{1}{2^{2.78}} + \frac{1}{3^{2.78}} + \frac{1}{4^{2.78}} + \frac{1}{5^{2.78}} + \dots + \frac{1}{10^{2.78}}$$

$$\sum_{x=1}^{P-1} \frac{1}{x^{2.78}} = \frac{1}{1} + \frac{1}{6.8685} + \frac{1}{21.203} + \frac{1}{47.177} + \frac{1}{87.728} + \dots + \frac{1}{602.56}$$

$$\sum_{x=1}^{P-1} \frac{1}{x^{2.78}} = 1.2437$$

b) Segunda equação

$$\frac{1}{(2.78-1)(11)^{2.78-1}} = \frac{1}{1.78(71.397)} = \frac{1}{127.09} = 0.007868$$

c) Terceira equação

$$\frac{1}{2(11)^{2.78}} = \frac{1}{1570.7} = 0.00064$$

d) Quarta equação

$$\frac{2.78}{24(11-1)^{2.78+1}} = \frac{2.78}{24(10)^{3.78}} = \frac{2.78}{24(6025.6)} = \frac{2.78}{1446144} = 0.000019$$

e) Por último, somando os valores:

$$1.2437 + 0.007868 + 0.00064 + 0.000019 = 1.252227$$

f) Finalmente,

$$C = \frac{1}{1.252227} = 0.798577$$

Arrendondando a quatro cifras

$$C = 0.7986$$

### 6.4 Cálculo dos valores esperados ou teóricos

Com os valores dos parâmetros  $n = -2.78$  e  $C = 0.7986$ , já conhecidos, calcular as frequências esperadas ou teóricas usando a equação (1)

$$y_x = C x^{-n}, \quad x = 1, 2, \dots, \max \quad (1)$$

Esta equação é uma simples multiplicação da constante C pelo número de contribuições em artigos  $x$ , elevados ao valor do expoente negativo de  $n$ , de modo que, para resolvê-lo, a equação (1) converte-se em:

$$y_x = C \left( \frac{1}{x^n} \right)$$

Depois, usando esta equação, calcule os valores esperados ou teóricos da distribuição de frequências.

a) Para  $x = 1$  (o número de autores que produziram 1 artigo)

$$y_1 = 0.7986 \times \frac{1}{1^{2.78}} = 0.7986 \times 1 = 0.7986$$

$$\text{Agora, } 0.7986 \cdot 962 = 768.2532 \pm 768.3$$

b) Para  $x = 2$  (o número de autores que produziram 2 artigos)

$$y_2 = 0.7986 \times \frac{1}{2^{2.78}} = 0.7986 \times 0.207 = 0.165$$

Agora,  $0.1163 \cdot 962 = 111.8806 \pm 111.9$

c) Para  $x = 3$  (o número de autores que produziram 3 artigos)

$$y_3 = 0.7986 \times \frac{1}{3^{2.78}} = 0.7986 \times \frac{1}{21.203} = 0.0377$$

Agora,  $0.0377 \cdot 962 = 36.2674 \pm 36.3$

d) Para  $x = 4$  (o número de autores que produziram 4 artigos)

$$y_4 = 0.7986 \times \frac{1}{4^{2.78}} = 0.7986 \times \frac{1}{47.177} = 0.01693$$

Agora,  $0.01693 \cdot 962 = 16.28666 \pm 16.3$

y assim sucessivamente ate atingir os valores de  $x = 11$

e) Para  $x = 11$  (o número de autores que produziram 11 artigos)

$$y_{11} = 0.7986 \times \frac{1}{11^{2.78}} = 0.7986 \times \frac{1}{785.37} = 0.001016846$$

Agora,  $0.001017 \cdot 962 = 0.9782 \pm 1.0$

## 6.5 Estabelecer as hipóteses

O que se vai testar é se a distribuição do poder inverso generalizado, obtida experimentalmente pelo método dos mínimos quadrados lineares, é homogênea ou não. Ou seja, se a probabilidade de que um elemento incluído na amostra é a mesma (igualmente provável) para todos os elementos nessa mesma situação. Portanto, estabelecemos as hipóteses da seguinte maneira:

$H_0$  = a distribuição representa a contagem de  $x = 1, 2, 3, \dots$  artigos por  $y$  autores.

$H_a$  <sup>1</sup> a distribuição não representa a contagem de  $x = 1, 2, 3, \dots$  artigos por  $y$  autores.

## 6.6 Especificar a região de rejeição das hipóteses ao nível de significância de $\alpha = .01$

Usando o nível de significância de  $\alpha = 0.01$ , na tabela dos valores críticos do teste Kolmogorov-Smirnov (K-S) de qualquer texto estatístico (ver

Anexo), encontrar a região de rejeição. Observando nessa tabela a coluna de  $n = 962$  ( $n$  é o tamanho da amostra, portanto, neste caso  $n > 40$ ) e na coluna do nível de significância de  $\alpha = 0.01$ , verificamos que o valor crítico do desvio máximo deve ser calculado, utilizando a fórmula seguinte:

$$\frac{1.63}{\sqrt{n}}$$

Isto significa dividir 1.63 pela raiz quadrada da população total de autores em estudo e listado na tabela da distribuição, isto é:

$$\frac{1.63}{\sqrt{962}} = \frac{1.63}{31.0161} = 0.05255$$

Portanto, o valor crítico do teste Kolmogorov-Smirnov (K-S) é igual a 0.05255

## 6.7 Teste de ajuste Kolmogorov-Smirnov da distribuição teórica da produtividade de autores

O teste de ajuste Kolmogorov-Smirnov (K-S) é um simples método não-paramétrico de testar se existem diferenças significativas entre as frequências observadas e as frequências teóricas ou calculadas de uma distribuição. É mais uma outra medida da bondade do ajuste de uma distribuição de frequências similar ao  $\chi^2$  (qui-quadrado). Não obstante, este teste Kolmogorov-Smirnov (K-S) é mais poderoso que o  $\chi^2$  (qui-quadrado), mais fácil de usar, e não precisa que os dados estejam agrupados em frequências inferiores a 5 como exige o teste  $\chi^2$  (qui-quadrado). É particularmente útil para julgar o grau de proximidade das frequências observadas face as frequências calculadas ou esperadas.

Com os valores de  $n$  e  $c$ , obtidos nas etapas 3 e 4, construir outra tabela com sete colunas. As colunas 1 e 2 contém os valores de  $x$  e  $y$  respectivamente. A coluna 3 contém a porcentagem de autores fazendo 1, 2, 3, 4, e assim em diante, contribuições nos dados observados. A coluna 4, contém os valores acumulados da coluna 3. A coluna 5, contém os valores teóricos relativos computados, usando  $n$  e  $C$  na formula  $C(1/x^n)$ . A coluna 6, contém os valores acumulados da coluna 5. A coluna 7 contém a diferença entre os pares de valores da coluna 4 e a coluna 6 (coluna 4 menos coluna 6). Na

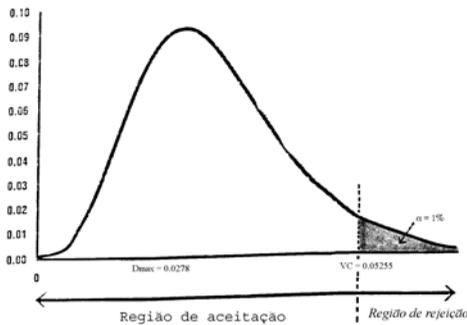
coluna 7 identifique a desviação máxima ( $D_{max}$ ). Este é o máximo valor absoluto encontrado nas diferenças dos valores das colunas 4 e 6. Valor absoluto significa tomar o valor calculado sem

considerar os signos negativos da desviação máxima calculada. No caso do trabalho de Estabrooks; Winther e Derksen (2004), a  $D_{max}$  é igual a 0.0278, ressaltado em negrito na Tabela 3 seguinte.

**Tabela 3:** Teste de ajuste Kolmogorov-Smirnov (K-S)

1	2	3	4	5	6	7
$x$	$y$	$\frac{y_x}{\sum y_x}$	$\sum \left[ \frac{y_x}{\sum y_x} \right]$	$c\left(\frac{1}{x^n}\right)$	$\sum \left[ c\left(\frac{1}{x^n}\right) \right]$	$D_{max}$
1	795	0.8264	0.8264	0.7986	0.7986	<b>0.0278</b>
2	103	0.1071	0.9335	0.1163	0.9149	0.0186
3	33	0.0343	0.9678	0.0377	0.9526	0.0152
4	12	0.0125	0.9802	0.0169	0.9695	0.0107
5	6	0.0062	0.9865	0.0091	0.9786	0.0079
6	5	0.0052	0.9917	0.0055	0.9841	0.0076
7	2	0.0021	0.9938	0.0036	0.9877	0.0061
8	3	0.0031	0.9969	0.0025	0.9902	0.0067
9	2	0.0021	0.9990	0.0018	0.9920	0.0070
10	0	0.0000	0.9990	0.0013	0.9933	0.0057
11	1	0.0010	1.0000	0.0010	0.9943	0.0057

Depois, compare o tamanho do valor crítico encontrado com o valor da  $D_{max}$ , de modo que agora só existam duas alternativas: maior ou menor. É o valor crítico de 0.05255, maior ou menor que a  $D_{max}$  0.0278? Neste caso, o valor crítico de 0.05255 é maior que a  $D_{max}$  de 0.0278. Portanto, a região de rejeição das hipóteses é a parte pontilhada e obscurecida da seguinte Figura 3. A região de aceitação é a parte à esquerda dessa figura.



**Figura 3:** Gráfico da região de aceitação e rejeição das hipóteses

### 6.8 Interpretação do ajuste da distribuição à lei de Lotka

Já que se tem apenas duas alternativas, existe somente duas formas de interpretação: a distribuição se ajusta à Lei de Lotka ou a distribuição não se ajusta à Lei de Lotka.

- a) Se a  $D_{max}$  é maior que o valor crítico (i.e., o valor crítico é menor que a  $D_{max}$ ), rejeite a hipótese nula de homogeneidade da distribuição de freqüências dos autores produtores de literatura de enfermagem, ou seja, rejeite a hipótese de que esta distribuição se ajusta à Lei de Lotka, ao 0.01 nível de significância.
- b) Se a  $D_{max}$  é menor que o valor crítico (i.e., o valor crítico é maior que a  $D_{max}$ ), aceite a hipótese nula de homogeneidade da distribuição de freqüências dos autores produtores de literatura de enfermagem, ou seja, aceite a hipótese de que esta distribuição se ajusta à Lei de Lotka, ao 0.01 nível de significância.

Como neste caso o valor crítico 0.05255 é maior que a  $D_{max}$  0.0278, então, esta distribuição

se ajusta à Lei de Lotka, a um 0.01 nível de significância.

### 6.9 Comparação dos valores observados e os valores calculados

Elabore uma tabela especial para comparar os valores observados e

calculados. Observe nesta tabela quão próximos ou afastados estão ambos os valores. Observe também que os totais são os mesmos, quase os mesmos, ou divergentes. A Tabela 4 mostra os valores observados e esperados para o caso da literatura de enfermagem, estudada por Estabrooks, Winther e Derksen (2004).

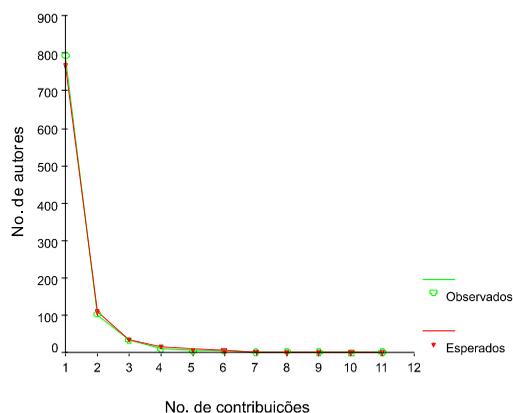
**Tabela 4:** Comparação dos valores observados e esperados

Contri- buições por autor x	Freqüência observada de autores fo	Freqüência esperada de autores fe	Divergencias
1	795	768.3	+ 26.7
2	103	111.9	- 8.9
3	33	36.3	- 3.3
4	12	16.3	- 4.3
5	6	8.8	- 2.8
6	5	5.3	- 0.3
7	2	3.4	- 1.4
8	3	2.3	+ 0.7
9	2	1.7	+ 0.3
10	0	1.3	- 1.3
11	1	1.0	0.0
Total	962	956.6	+ 5.4

Nesta tabela, as maiores divergencias se apresentam nas primeiras frequencias observadas da distribuição dos autores produtores de artigos. O modelo estimou 26.7 autores menos do que os 795 autores observados como produtores de um único artigo, bem como 8.9 autores a mais dos 103 autores observados como produtores de dois artigos. Estas divergencias parecem incidir também na estimação do total de autores produtores, pois existe uma diferença de 5.4 autores menos do que os 962 autores observados.

### 6.10 Gráfico de dispersão dos valores observados e esperados

Como se mostra na Figura 4, a aproximação entre os valores observados e esperados da distribuição da produtividade de autores pode ser melhor observada do traçado da dispersão de ambos os valores. Esta figura deve ser incluída na redação do informe final.



**Figura 4:** Dispersão dos valores observados e esperados

### 6.11 Interpretação do ajuste da distribuição à lei de Lotka

Já que se tem apenas duas alternativas, existe somente duas formas de interpretação: a

distribuição se ajusta à Lei de Lotka ou a distribuição não se ajusta à Lei de Lotka.

- a) Se a  $D_{max}$  é maior que o valor crítico (i.e., o valor crítico é menor que a  $D_{max}$ ), rejeite a hipótese nula de homogeneidade da distribuição de frequências dos autores produtores de literatura de enfermagem, ou seja, rejeite a hipótese de que esta distribuição se ajusta à Lei de Lotka, ao 0.01 nível de significância.
- b) Se a  $D_{max}$  é menor que o valor crítico (i.e., o valor crítico é maior que a  $D_{max}$ ), aceite a hipótese nula de homogeneidade da distribuição de frequências dos autores produtores de literatura de enfermagem, ou seja, aceite a hipótese de que esta distribuição se ajusta à Lei de Lotka, ao 0.01 nível de significância.

Como neste caso o valor crítico 0.05255 é maior que a  $D_{max}$  0.0278, então, esta distribuição se ajusta à Lei de Lotka, a um 0.01 nível de significância.

### 6.12 Comparação dos valores observados e os valores calculados

Elabore uma tabela especial para comparar os valores observados e calculados. Observe nesta tabela quão próximos ou afastados estão ambos os valores. Observe também que os totais são os mesmos, quase os mesmos, ou divergentes. A Tabela 4 mostra os valores observados e esperados para o caso da literatura de enfermagem, estudada por Estabrooks, Winther e Derksen (2004).

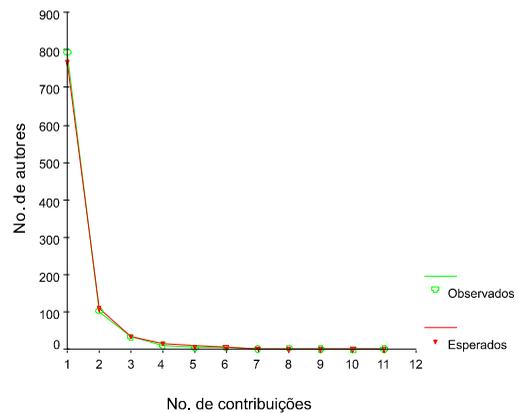
**Tabela 4:** Comparação dos valores observados e esperados

Contri- buições por autor x	Frequência observada de autores fo	Frequência esperada de autores fe	Divergencias
1	795	768.3	+ 26.7
2	103	111.9	- 8.9
3	33	36.3	- 3.3
4	12	16.3	- 4.3
5	6	8.8	- 2.8
6	5	5.3	- 0.3
7	2	3.4	- 1.4
8	3	2.3	+ 0.7
9	2	1.7	+ 0.3
10	0	1.3	- 1.3
11	1	1.0	0.0
Total	962	956.6	+ 5.4

Nesta tabela, as maiores divergencias se apresentam nas primeiras frequencias observadas da distribuição dos autores produtores de artigos. O modelo estimou 26.7 autores menos do que os 795 autores observados como produtores de um único artigo, bem como 8.9 autores a mais dos 103 autores observados como produtores de dois artigos. Estas divergencias parecem incidir também na estimação do total de autores produtores, pois existe uma diferença de 5.4 autores menos do que os 962 autores observados.

### 6.13 Gráfico de dispersão dos valores observados e esperados

Como se mostra na Figura 4, a aproximação entre os valores observados e esperados da distribuição da produtividade de autores pode ser melhor observada do traçado da dispersão de ambos os valores. Esta figura deve ser incluída na redação do informe final.



**Figura 4:** Dispersão dos valores observados e esperados

## 7 CONCLUSÕES

Foram encontrados 962 autores que, conjuntamente, produziram 1275 artigos, sendo que 83% deles contribuíram com um único artigo à literatura estudada. O modelo permitiu também identificar uma alta porcentagem de 97.9% pequenos produtores, responsáveis por 90.2% da literatura publicada. Os medianos e grandes produtores da literatura de enfermagem corresponderam a 2.1% dos autores responsáveis por 9.8% da literatura produzida.

O modelo do poder inverso generalizado pelo método dos mínimos quadrados e o teste

Kolmogorov-Smirnov (K-S) foram usados para avaliar o ajuste dos dados observados e esperados. Com  $n = -2.78$  e  $C = 0.7986$ , ao 0.01 nível de significância, verificou-se que o valor crítico foi de 0.05255, com uma desviação máxima de 0.0278, portanto, esta literatura ajusta-se muito bem ao modelo de Lotka. Esses resultados são diferentes dos reportados por Estabrooks, Winther e Derksen

(2004). Embora esses autores afirmam que a distribuição da produtividade dos autores de literatura no campo da enfermagem foi "consistente com a lei de Lotka" não estimaram o valor do expoente  $n$  nem testaram estatisticamente seus dados observados. Esses fatores podem ser as causas das divergências encontradas na replica realizada neste trabalho.

## THE PRODUCTIVITY OF NURSING LITERATURE 'S AUTHOURS: an application of Lotka 's Law Model

### ABSTRACT

This work describes the nature of the generalized inverse power distribution by the minimum linear squared method. It describes, step by step, the process of application of this model using the data, collected and studied by Estabrooks; Winther & Derksen on nursing literature . With  $n = - 2.78$ , and  $C = 0.7986$ , at 0.01 level of significance, a critical value of 0.05255 much higher than the estimated maximum deviation of 0.0278, was found. Therefore, this literature distribution fits Lotka's model.

### Keywords

LOTKA'S LAW  
GENERALIZED INVERSE POWER DISTRIBUTION  
AUTHOR'S PRODUCTIVITY  
BIBLIOMETRICS  
NURSING  
INFOMETRICS  
SCIENTOMETRICS

---

Artigo recebido em 23.09.2005 e aceito para publicação em 19.06.2006

---

### REFERÊNCIAS

ALBEROLA, A. et.al.. Estudio bibliométrico de "Arbor". *Arbor: ciencia, pensamiento y cultura*, Madrid, 122(479-80):125-137, nov.-dic. 1985.

AL-GHAMDI, Ali; et. al. *Authorship in JASIS: a quantitative analysis*. Katharine Sharp Review, no. 6, Winter 1998. ([http://www.lis.uiuc.edu/review/6/al\\_ghamdi.html](http://www.lis.uiuc.edu/review/6/al_ghamdi.html)).

BUDD, John M. A bibliometric analysis of higher education literature. *Research in Higher Education*, v.28, n.2, p.180-190, 1988.

CARR, Lowell Juilliard. The patenting performance of 1,000 inventors during ten years. *The American Journal of Sociology*, v.37, n.4, p.569-580, jan. 1932.

CATTELL, J. M. A further statistical study of American men of science. *Science*, v.32, n.633-648, p.672-688, 1910.

CHUNG, Kee H.; COX, Raymond A. K. Patterns of productivity in the finance literature: a study of the bibliometric distributions. *The Journal of Finance*, v.45, n.1, p.301-309, Mar. 1990.

CHUNG, Kee H.; PAK, Hong S.; COX, Raymond A. K. Patterns of research output in the accounting literature: a study of the bibliometric distributions. *Abacus*, v.28, n.2, p.168-185, 1992.

CLINE, Gloria S. College & Research Libraries: its first forty years. *College & Research Libraries*, v.43, n.3, p.208-232, may 1982.

COOK, Kevin L. Laws of scattering applied to popular music. *Journal of the American Society*

- for *Information Science*, v.40, n.4, p.277-283, Jul. 1989.
- COX, Raymond A. K.; CHUNG, Kee H. Patterns of research output and author concentration in the economics literature. *The Review of Economics and Statistics*, v.73, n.4, p.740-747, Nov. 1991.
- COX, Raymond A. K.; FELTON, James M.; CHUNG, Kee H. The concentration of commercial success in popular music: an analysis of the distribution of gold records. *Journal of Cultural Economics*, v.19, n.4, p.333-340, 1995.
- Dennis, Wayne. Productivity among American psychologists. *American Psychologist*, v.9, n.5, p.191-194, May. 1954.
- DENNIS, Wayne. Variations in productivity among creative workers. *The Scientific Monthly*, v.80, n.4, p.277-278, Apr. 1955.
- DRESDEN, A. A report on the scientific work of the Chicago Section, 1897-1922. *Bulletin of The American Mathematical Society*, v.28, p.303-307, Jul. 1922.
- DUFRENOY, Jean. The publishing behavior of biologists. *Quarterly Review of Biology*, v.13, p.207-210, 1938.
- ESTABROOKS, Carole A.; WINTHER, Connie; DERKSEN, Linda. Mapping thwe field: a bibliometric analysis of the research utilization literature in nursing. *Nursing Research*, v.53, n.5, p.293-303, Sept.-Oct. 2004.
- FOX, Wendy R.; LASKER, Gabriel W. The distribution of surname frequencies. *International Statistical Review*, v.51, n.1, p.81-87, Apr. 1983.
- GUPTA, B. M.; KUMAR, Suresh; ROUSSEAU, R. Applicability of selected probability distributions to the number of authors per article in theoretical population genetics. *Scientometrics*, v.42, n.3, p.325-334, 1998.
- GUPTA, B. M. et.al. Distribution of productivity among authors in potato research (1900-1980). *Library Science with a Slant to Documentation*, v.33, n.3, p.127-134, Sept. 1996.
- GUPTA, Davendra K. Lotka's law and productivity patterns of entomological research in Nigeria for the period, 1900-1973. *Scientometrics*, v.12, n.1-2, p.33-46, Jul. 1987.
- HERSH, A. H. Drosophila and the course of research. *Ohio Journal of Science*, v.42, n.5, p.198-200, 1942.
- IVANOV, I. V.; LUKOVSKAYA, T. S. Scientometric (Bibliometric) analysis of publications in *Prochvovedenie* for 100 years (1899-1998). *Eurasian Soil Science*, v.36, n.1, p.107-120, 2003.
- JIMÉNEZ CONTRERAS, Evaristo; MOYA DE ANEGÓN, Félix. Análisis de la autoría en revistas españolas de Biblioteconomía y Documentación, 1975-1995. *Revista Española de Documentación Científica*, v.20, n.3, p.252-266, 1997.
- LEAVENS, Dickson H. Letter to the Editor. *Econometrica*, v.21, n.4, p.630-632, Oct. 1953.
- LÓPEZ CALAFI, J.; SALVADOR, A.; GUARDIA, M. de la. Estudio bibliométrico de la literatura sobre la determinación de elementos metálicos en aceites lubricantes por espectroscopía atómica. *Revista Española de Documentación Científica*, v.8, n.3, p.201-213, 1985.
- LÓPEZ-MUÑOZ, Francisco; RUBIO VALLADOLID, G. La producción científica española en psiquiatría: un estudio bibliométrico de las publicaciones de circulación internacional durante el periodo 1980-1983. *Anales de Psiquiatría*, v.2, n.2, p.68-75, 1995.
- LÓPEZ-MUÑOZ, Francisco et.al. Scientific research on the pineal gland and melatonin: a bibliometric study for the period 1966-1994. *Journal of Pineal Research*, v.20, n.3, p.115-124, 1996.
- LOTKA, Alfred J. The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, v.16, n.12, p.317-323, June 19, 1926.
- LOUGHNER, William. Lotka's law and the Kolmogorov-Smirnov test: an error in calculation. *Journal of the American Society for Information Science*, v.43, n.2, p.149-150, March 1992.
- NATH, Ravinder; JACKSON, Wade M. Productivity of management information systems researchers: does Lotka's law applied? *Information Processing & Management*, v.27, n.2/3, p.203-209, 1991.
- OLIVEIRA, Margarida Pinto; CALDEIRA, Paulo da Terra. Análise bibliométrica da literatura médica brasileira. *Revista da Escola de Biblioteconomía da Universidade Federal de Minas Gerais*, v.5, n.1, p.7-26, Mar. 1976.

- OPPENHEIM, C. Use of online databases in bibliometric studies. In: *INTERNATIONAL ONLINE INFORMATION MEETING, 9th, London 3-5 december 1985*. Oxford : *Learned Information*, 1986. p. 355-364.
- PAO, Miranda Lee. An empirical examination of Lotka's law. *Journal of the American Society for Information Science*, v.37, n.1, p.26-33, Jan. 1986.
- POTTER, William Gray. Lotka's law revisited. *Library Trends*, v.30, n.1, p.21-39, Summer 1981.
- SÁNCHEZ HERNÁNDEZ, Antonio; QUIÑONES VIDAL, Elena; CAMEO OLCINA, Mercedes. La revista Anales de Psicología desde un punto de vista historiográfico: La primera década de una publicación con raíces académicas (1984-1994). *Anales de Psicología*, v.11, n.1, p.1-18, 1995.
- URBIZAGÁSTEGUI ALVARADO, Rubén. A lei de Lotka na bibliometria brasileira. *Ciência da Informação*, Brasília, v.31, n.2, p.14-20, maio/ago. 2002.
- URBIZAGÁSTEGUI ALVARADO, Rubén. La ley de Lotka y la literatura de Bibliometría. *Investigación Bibliotecológica*, v.13, n.27, p.125-141. 1999.
- URBIZAGÁSTEGUI ALVARADO, Rubén. *Lotka's law on Lotka's literature: an exploration*. Riverside, Calif, 2002. Datilografado.
- URBIZAGÁSTEGUI ALVARADO, Rubén; OLIVEIRA, Marlene de. A produtividade dos autores na antropologia brasileira. *DataGramZero*, v.2, n.6, p.1-17, Dez. 2001.
- URBIZAGÁSTEGUI ALVARADO, Rubén; CORTÉS, María Teresa. La productividad de autores en la Revista Geológica de Chile. *Ciencia de la Información*, Cuba, v.33, n.2, p.24-36, 2002.
- VLACHÝ, Jan. Evaluating the distribution of individual performance. *Scientia Yugoslavica*, v.6, n.1/4, p.267-275, 1980.
- WAGNER-DOBLER, Roland; BERG, Jan. Nineteenth-century mathematics in the mirror of its literature: a quantitative approach. *Historia Mathematica*, v.23, n.3, p.288-318, Aug. 1996.
- WILLIAMS, C. B. The number of publications written by biologists. *Annals of Eugenics*, v.12, p.143-146, 1944.

**APÊNDICE**

**Tabela I:** Valores críticos do teste Kolmogorov-Smirnov

n	α = 0,10		α = 0,05		α = 0,01	
	D	D <sub>lim</sub>	D	D <sub>lim</sub>	D	D <sub>lim</sub>
10	0,200	0,230	0,173	0,196	0,108	0,143
15	0,173	0,196	0,146	0,169	0,091	0,126
20	0,146	0,169	0,119	0,142	0,074	0,109
25	0,119	0,142	0,092	0,115	0,057	0,092
30	0,092	0,115	0,065	0,088	0,040	0,075
35	0,065	0,088	0,038	0,061	0,023	0,058
40	0,038	0,061	0,016	0,039	0,006	0,041
45	0,016	0,039	0,009	0,032	0,000	0,034
50	0,009	0,032	0,003	0,025	0,000	0,027
55	0,003	0,025	0,000	0,018	0,000	0,020
60	0,000	0,018	0,000	0,011	0,000	0,013
65	0,000	0,011	0,000	0,004	0,000	0,006
70	0,000	0,004	0,000	0,000	0,000	0,002
75	0,000	0,000	0,000	0,000	0,000	0,000
80	0,000	0,000	0,000	0,000	0,000	0,000
85	0,000	0,000	0,000	0,000	0,000	0,000
90	0,000	0,000	0,000	0,000	0,000	0,000
95	0,000	0,000	0,000	0,000	0,000	0,000
100	0,000	0,000	0,000	0,000	0,000	0,000