Biblios

ISSN 1562-4730 (online)

Alguns métodos estatísticos voltados às unidades de informação

Adilson Luiz Pinto
Universidade Federal de Santa Catarina - UFSC. Brasil

Alexandre Oliveira de Meira Gusmão,
André de Souza Pena
Universidade Federal de Mato Grosso - UFMG. Brasil

Marcelo Moreira Ferreira da Silva
Universidade Federal de Juiz de Fora - UFJF. Brasil

ARTIGOS / ARTICLE

Resumo

Este estudo visa mostrar algumas técnicas estatísticas aplicadas na gestão informacional nas unidades de informação através da utilização de métodos de mensuração e métricas quantitativas atreladas a estudos relacionados à bibliometria, cienciometria, econometria entre outros. Os estudos envolvendo estatística podem ser descritivos, para construção de indicadores, inferencial através de modelos teóricos ou empíricos. As técnicas estatísticas permitem que de base de dados sejam extraídas informações relevantes à tomada de decisão. Assim, será brevemente explanando sobre probabilidade, distribuição de frequência, séries estatísticas, covariância e correlação, regressão linear e números índices.

Palavras-chave
Estatística aplicada; Unidades de informação; Estudos métricos; Teorías métricas

Some statistical methods directed the information centers

Abstract

This study aims to show some statistical techniques applied in information management in the units of information through the use of methods of measurement and quantitative metrics tied to studies related to bibliometrics, scientometrics, econometrics among others. The studies involving statistics can be descriptive, for construction of indicators, inferential models through theoretical or empirical. Statistical techniques allow the database to be extracted relevant information for decision making. This will briefly explaining about probability, frequency distribution, statistical series, covariance and correlation, linear regression and index numbers.

Keywords
Applied statistics; Information centers; Metric studies; Metric theories

1. Considerações iniciais

A Estatística aplicada à Biblioteconomia e à Documentação é utilizada no controle dos produtos e serviços de informação e para melhor determinar, explicar, compreender e apresentar soluções relacionadas aos sistemas e fluxos de informação, intervindo no processo de planejamento, execução e controle. Independente da aplicação a Estatística é classificada em:

- Estatística Descritiva – que procura descrever e analisar um determinado fenômeno e reúne os passos iniciais do processo estatístico, que são a coleta, a organização, a descrição e apresentação dos dados, também conhecida como síntese dos dados e;
- Estatística Dedutiva ou Inferencial – Para além de descrever os dados, a estatística, através de seu ferramental, busca ajudar a inferir até que pontos os resultados advindos de uma amostra são representativos da população em estudo. A estatística inferencial é responsável por tratar das condições em que as inferências, apresentadas nas conclusões podem ser aceitas como válidas. As inferências (extensão das conclusões baseadas em uma amostra que é generalizada para a totalidade da população) não são absolutamente corretas, estão associadas a um grau de incerteza provindo de variações aleatórias dos dados amostrais, por isso faz-se a utilização da linguagem das probabilidades para o estabelecimento de diretrizes que possibilitem a validade das conclusões.

Entre os conceitos mais importantes da estatística destacamos, dentro das unidades de informação, alguns como:

- População ou Universo: é o conjunto dos elementos que possuem pelo menos uma característica em comum, que está inserida no contexto da análise, e do qual se deseja extrair uma informação. No caso específico da Biblioteconomia a população pode ser todos os usuários de uma determinada unidade de informação.
- Censo: é o estudo estatístico que envolve todos os elementos da população de interesse, ou seja, quando se observam todos os elementos da população estatística, como todos os alunos e professores em uma biblioteca e que estão cadastrados no sistema como usuários.
- Amostra: O espaço amostral é um conjunto pertencente ao universo formado por todos os resultados possíveis que podem ser amostrados ou observados, assim podemos definir amostra como um subconjunto dos elementos da população, ou uma partição do espaço amostral, que tem por missão representar o conjunto inteiro. Costumam-se tomar amostras quando a população é grande, quando se deseja o resultado da pesquisa em curto espaço de tempo ou quando se deseja conter gastos. Para exemplificar suponha-se uma universidade que tenha várias bibliotecas e que o espaço amostral seja formado por todos os usuários de todas elas. A amostra poderá ficar restrita a uma só biblioteca ou a um grupo reduzido de usuários selecionados por algum critério.
- Indivíduos, unidades estatísticas ou caso: Chama-se unidade estatística, caso ou indivíduo a cada um dos elementos que compõem a população estatística. O indivíduo é um dado observável que pode ser uma pessoa, um objeto ou inclusive um dado abstrato. No caso de uma biblioteca ou unidade de informação, podem ser os usuários tratados particularmente.
- Variáveis ou atributos: São conjunto de características, qualidades ou propriedades inerente aos indivíduos, por exemplo, idade, altura, sexo, marca, cor, dentre outros. No contexto estatístico, uma variável pode tanto ser determinística, que é o caso de uma medida realizada, ou aleatória, na qual não se tem certeza de sua medida exata. As variáveis aleatórias são representadas por modelos probabilísticos. Um indivíduo ou casso pode possuir um ou mais atributos ou variáveis os quais são tratados como variáveis aleatórias caso ainda não tenham sido medidos. A variável é o objeto do estudo que se converte em um atributo mensurável e seus valores podem variar entre indivíduos assumindo determinadas características dentro de uma pesquisa. As variáveis classificam-se em qualitativas e quantitativas e em independentes, dependentes e intervenientes.

Os tipos de variáveis são: Qualitativas, que são categóricos, mas não são numéricos e Quantitativos, que são cardinais.

As Variáveis Quantitativas são aquelas passíveis de se atribuir um valor numérico, representam quantidades e dividem-se em (1) Discretas, que tomam valores isolados, ou seja, os valores se apresentam em “saltos” e que não podem tomar nenhum valor intermediário entre dois números consecutivos fixados e (2) Contínuas, que podem

tomar infinitos valores dentro de um intervalo determinado de resultados possíveis, os pontos no intervalo formado por números contínuos são considerados pontos de acumulação, porque por menor que seja a distancia entre o ponto e um de seus vizinhos sempre haverá um vizinho ainda mais próximo, esse fato não acontece no intervalo formado por pontos isolados que é o caso dos intervalos discretos; e Variáveis Qualitativas que são aquelas para a qual não se pode atribuir um valor numérico no sentido de representar quantidades, estas se dividem em (1) Ordinais, que admitem uma ordenação ou hierarquização entre as respostas, ainda que seus resultados não sejam numéricos e (2) Nominais as que não admitem uma ordenação ou hierarquização entre as respostas (FERNÁNDEZ PALACÍN et al., 2000).

2. Probabilidade

Dentro das aplicações estatísticas, outro estudo empregado na área de Biblioteconomia e Documentação é o das probabilidades. A teoria da probabilidade é o ramo da matemática que estuda os fenômenos aleatórios. A noção intuitiva de probabilidade é a seguinte: seja **A** um resultado de um determinado experimento realizado **n** vezes. Assim, diremos que $\mathbf{n_A}$ é o número de vezes em que se obteve o resultado **A**. Desse modo, temos que $0 \leq \mathbf{n_A} \leq n$. Se a razão $\mathbf{n_A/n}$ se aproxima de um número **p** quando **n** fica indefinidamente grande, diremos que **p** é a probabilidade de ocorrências e escrevemos **p = P(A**). Além disso, pela definição, $0 \leq P(A) \leq 1$. Por exemplo, se o experimento consiste no lançamento de um dado e o evento **A** é "sair o número 5", então P(A) = 1/6, pois numa longa série de lançamentos a frequência com que o número 5 se aproxima de 1/6. Do mesmo modo, se o experimento for lançar uma moeda e **A** denotar "cara", então P(A) = 1/2. A probabilidade de um evento correto é 1; então se pode dizer que "domingo vem depois do sábado" com probabilidade igual a 1. Analogamente, a probabilidade de um evento impossível é zero; então podemos dizer que a probabilidade de obter o número 100 no lançamento de um dado é 0 (PEÑA; ROMO, 1997).

Freqüentemente o que se quer determinar é a probabilidade de certo resultado pertencente a algum conjunto. Por exemplo, pode-se estar interessado em determinar a probabilidade de um número par ao lançar um dado. Neste caso, o conjunto de resultados é hipoteticamente 2, 4 e 6. Um conjunto **E** deste tipo é dito um evento; se o resultado do experimento está em **E**, dizemos que o evento ocorreu. A probabilidade de um evento **E** é definida da mesma forma que no caso de um resultado simples; repete-se o experimento **n** vezes, denotamos $\mathbf{n_E}$ como a representação numérica de vezes que **E** ocorreu e definimos **P(E)** como o número ao qual $\mathbf{n_E/n}$ se aproxima quando **n** é indefinidamente grande. No exemplo anterior, **E = {2, 4, 6}**, temos que **P(E) = 1/2**.

Então, **A** e **B** são dois eventos. Denota-se por **A**U**B** o evento obtido ao que combinamos os resultados da **A** e **B** num único conjunto. Segue pela definição de probabilidade que se **A** e **B** são adjuntos, isto é, não têm resultados em comum, então: P(AUB) = P(A) + P(B)

Por exemplo, no experimento que consiste em atirar um dado, sejam A = {5} e B = 2, 4, 6. Então A e B são adjuntos, AUB = {2, 4, 5, 6} e P(AUB) = 1/6 + 1/2 = 2/3.

No caso de atirar um dado ou uma moeda, é bastante intuitivo que não é necessário realizar o experimento para determinar as probabilidades de vários eventos. Isso se deve, nesses casos, à simetria entre as caras de uma moeda ou os lados de um dado. Se um experimento com um número finito **N** de possíveis resultados e se por alguma razão podermos considerar que eles são igualmente prováveis, então podemos associar a cada um deles uma probabilidade igual a **1/N**. Assim, **N = 6** para um dado, onde cada um tem probabilidade de 1/6 (PEÑA; ROMO, 1997).

Para a representação matemática, a probabilidade de um evento **E** é então definida como $\mathbf{N_E/N}$, onde $\mathbf{N_E}$ é o número de resultados no conjunto **E**. Quando se calcula essa quantidade, é conveniente referir aos resultados em **E** como os

que são favoráveis e os que não são favoráveis. Assim, pode-se dizer que a probabilidade do evento **E** é: **P ($_E$)** = número de resultados favoráveis ÷ número total de resultados. No entanto, enfatiza-se que isso é verdadeiro somente no caso em que **N** resultados sejam igualmente prováveis.

Para entender as relações entre as variáveis quantitativas, existe a técnica estatística chamada Análise de Regressão, na qual são relacionadas por uma equação duas variáveis: uma chamada variável resposta ou dependente, e outra, chamada variável explicativa, ou independente. Neste caso, trata-se de Regressão Linear simples, pois envolve apenas duas variáveis. Para se estimar o valor esperado, usa-se de uma equação, que determina a relação entre ambas as variáveis:

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

Em que: Yi - Variável explicada (dependente); é o valor que se quer atingir;

α - É uma constante, que representa a interceptação da reta com o eixo vertical;

β - É outra constante, que representa o declive da reta;

X_i - Variável explicativa (independente) representa o fator explicativo na equação;

ε_i - Variável que inclui todos os fatores residuais mais os possíveis erros de medição. O seu comportamento é aleatório, devido à natureza dos factores que encerra. Para que essa fórmula possa ser aplicada, os erros devem satisfazer determinadas hipóteses, que são: serem variáveis normais, com a mesma variância σ^2 (desconhecida), independentes e independentes da variável explicativa *X*.

As Variáveis Independentes são aquelas que afetam, influenciam ou determinam outra variável, também podem ser consideradas como a condição ou a causa, para um determinado efeito ou conseqüência. As variáveis independentes podem ser representadas pela função ***y = f(x)***, tal que para qualquer real ***x*** (variável independente) tem-se uma variável dependente ***y***.

As Variáveis Dependentes são aqueles efeitos, resultados, conseqüências ou resultados afetados ou explicados pela variável independente e variará conforme as mudanças na variável independente, ou seja, é o fator que varia à medida que o pesquisador modifica a variável independente. As Variáveis Intervenientes são aquelas que, "numa seqüência causal, se coloca entre a variável independente (I) e a variável dependente (D), tendo a função de ampliar, anular ou diminuir a influência de I sobre D" (LAKATOS; MARCONI, 2001, p. 150), porém não podem ser manipuladas ou medidas pelo pesquisador. Na continuação apresentam-se as séries estatísticas e as normas para apresentação dos dados estatísticos.

3. Séries estatísticas

As Séries Estatísticas são maneiras de apresentar os dados estatísticos de uma forma tabulada, ou seja, são todas e quaisquer tabelas que apresentem a distribuição de um conjunto de dados estatísticos em função do objeto do estudo (descrição do fato), do local e da época. Estes elementos deverão responder as seguintes perguntas: O quê? Onde? e Quando?

As Séries Estatísticas classificam-se em (1) séries históricas - são aquelas cujo elemento variável é a época, elas são denominadas também de cronológica, temporais ou de marcha; (2) séries geográficas – nestas séries o elemento variável é o local, elas são conhecidas também por séries espaciais, territoriais ou de localização; (3) séries específicas – são aquelas cujo elemento variável é a descrição do fenômeno; e (4) distribuição de frequência – são aquelas cujos dados são apresentados sob um critério de magnitude, em classes ou intervalos, permanecendo fixos os elementos fato, local e época.

Os dados, em geral, representam-se em forma de tabelas ou gráficos, técnica muito utilizada na área de Biblioteconomia e Documentação, assim é importante descrevermos estes tipos de visualização estatística.

As tabelas devem explicar na menor quantidade possível todo o contexto proposto, tem-se de dar suficiente informação no título e nos cabeçalhos das colunas e das linhas da tabela para permitir que o leitor identifique facilmente seu conteúdo, como fato, local, época, freqüência, bem como relações que se apresentem.

Segundo Babbie (2005) a regra principal é colocar a porcentagem para baixo e ler através, ou a porcentagem através e ler para baixo.

Os gráficos devem ser auto-explicativos e de fácil visualização. O conteúdo de um gráfico deve ser o mais completo possível. As escalas vertical e horizontal devem ser rotuladas com clareza para dar pertinência às unidades (ETHERINGTON, 2000). A maioria dos gráficos apresentam informação numérica com escalas, que devem rotular-se para descrever completamente a variável apresentada na escala. Para as variáveis de medida devem-se indicar as unidades de medição.

4. Distribuição de freqüências

A Distribuição de Freqüência é uma técnica estatística utilizada para apresentar uma coleção de dados classificados e agrupados em classes de modo a destacar a freqüência existente em cada classe. Para isto utiliza-se de uma tabela que associa a cada evento o número de vezes que ele ocorre, este número recebe o nome de freqüência. São elementos indispensáveis numa Tabela de Distribuição de Freqüência: (1) a Classe - agrupamento de valores em um determinado intervalo; (2) o Intervalo de Classe – os pontos extremos dentro de uma Classe; os Limites de uma Classe; o Ponto Médio de uma Classe; a Amplitude de uma Classe e; a Amplitude Total da Distribuição.

No universo estatístico existem 6 tipos de freqüência:

a) Freqüência Absoluta Simples (f_i): é o número de observações (n_i) correspondentes a determinada classe ou valor, e simbolizada por f_i (lê-se: f índice $_i$ ou freqüência da classe $_i$). Assim tem-se: $f_1 = n_1$, $f_2 = n_2$, $f_3 = n_3$, $f_i = n_i$. A f_i pode ser aplicada para identificar a quantidade total de livros por classe ou qualquer outra tipologia bibliográfica existente em uma unidade de informação.

b) Freqüência Absoluta Acumulada Crescente (***fac***): é a soma da freqüência anterior (f_1) com a freqüência posterior (f_2), realizada sucessivamente com todas as freqüências, considerando os valores acumulados, num movimento diagonal, e no sentido de cima para baixo da tabela de distribuição de freqüência. Assim temos: $fac_1 = f_1$, $fac_2 = f_1 + f_2$, $fac_3 = f_1 + f_2 + f_3$, $fac_i = f_1 + f_2 + f_3 + f_i$.

c) Freqüência Absoluta Acumulada Decrescente (***fad***): é a soma da última freqüência (***fni***) com a freqüência anterior (f_{ni-1}), realizada sucessivamente com todas as freqüências, considerando os valores acumulado, num movimento diagonal, e no sentido de baixo para cima em uma tabela de distribuição de freqüência. Assim temos: $fad_{ni} = f_{ni}$, $fad_{ni-1} = f_{ni} + f_{ni-1}$, $fad_{ni-2} = f_{ni} + f_{ni-1} + f_{ni-2}$.

d) Freqüência Relativa Simples: é a razão ou o coeficiente entre a frequência absoluta (f_i) da variável pelo total de observações (n_i) correspondentes a determinada classe ou valor, e simbolizada por (F_i).

e) Freqüência Relativa Acumulada Crescente (***Fac***): é razão da soma da primeira freqüência (f_1) com a freqüência posterior (f_2), e realizada sucessivamente.

f) Freqüência Relativa Acumulada Decrescente (***Fad***): é a razão da soma da última freqüência (f_{ni}) com a freqüência anterior (f_{ni-1}), e realizada sucessivamente com todas as freqüências (SÁNCHEZ CORONA, 2005).

As medidas típicas numa distribuição de freqüência são as medidas de posição, as medidas de dispersão e as medidas de assimetria e de curtose. As medidas de posição são medidas estatísticas que servem para orientar quanto à posição da distribuição em relação ao eixo horizontal do gráfico da curva de freqüência.

As medidas de posições mais importantes são as medidas de tendência central e as separatrizes. As medidas de tendência central são aquelas cujos valores tendem a localizarem-se no centro de uma série de dados. Freqüentemente, quando se analisa os valores de uma variável em uma amostra, constata-se que os dados não se distribuem uniformemente, havendo concentração em alguns pontos, notadamente próximos ao centro da distribuição. Ou seja, é comum haver um grande número de elementos com valores próximos à média e poucos apresentando valores extremos, isto é, próximos ao valor mínimo e máximo. As medidas de tendência central mais utilizadas são a média, a moda e a mediana. Outras menos usadas são as médias geométricas e a harmônica.

a) Média: é o valor que aponta para onde mais se concentram os dados de uma distribuição e pode ser considerada o ponto de equilíbrio das frequências, num histograma.
b) Mediana: é o ponto central de uma série de dados agrupados para os quais vem dada por uma interseção aproximada da centralidade.
c) Moda: é o valor que aparece com maior freqüência numa distribuição de freqüência.

As medidas separatrizes são aquelas medidas que separam ou que dividem o conjunto em certo número de partes iguais, e englobam: a própria mediana, os deciis, os quartiis e os percentiis.

As medidas de dispersão são as medidas (amplitude total, desvio quartílico, desvio médio, desvio padrão, variância, coeficiente de variação de Pearson, variância relativa) que indicam se os elementos de uma distribuição estão mais próximos ou afastados de um ponto de referência, que em geral é a própria média do conjunto (LEVIN; LEVIN, 1997).

a) Ranking: é a diferença entre o ponto mais alto e o mais baixo da distribuição. Exemplo: se a temperatura mais alta do Rio de Janeiro é 44°C e a mais fria é de 28°C, então a amplitude da temperatura anual desta cidade seria de 16° C (44-28 = 16). Dentro das unidades de informações os rankings são também utilizados para saber quais títulos são os mais buscados;

b) Desvio-padrão: é um valor que quantifica a dispersão ou espalhamento dos eventos sob distribuição normal em relação à média da distribuição, ou seja, a média das diferenças entre o valor de cada evento e a média central, encontrado calculando-se a raiz quadrada da Variância. Quanto maior o desvio-padrão, maior a dispersão e mais afastados da média estarão os eventos extremos.

c) Variância: é a média do quadrado da distância de cada ponto até a média, e indica quão longe em geral os valores observados se encontram do valor esperado.

d) Coeficiente de variação é uma medida de dispersão utilizada para comparar a variação de conjuntos de observações que diferem na média ou são medidos em grandezas diferentes. Quando o desvio-padrão de duas distribuições não é comparável, uma solução é usar o coeficiente de variação, que é igual ao desvio-padrão dividido pela média, pois os desvios-padrão só podem ser devidamente avaliados quando comparados sob a mesma grandeza.

O diagrama de dispersão é uma forma de visualização da estatística aplicada à Biblioteconomia e Documentação, que se representa como um ponto no espaço cartesiano XY, utilizados simultaneamente entre os valores de duas variáveis quantitativas (X, Y) medidas em cada elemento do conjunto de dados. O diagrama de dispersão é

utilizado para visualizar a relação e associação entre duas variáveis, mas também pode ser usado para visualizar a relação de dois tratamentos no mesmo indivíduo ou para verificar o efeito de uma amostra por determinação de uma análise anterior e de outra posterior do diagrama, conforme a seguinte representação:

Individuo	Variável X	Variável Y
A	2	3
B	4	3
C	4	5
D	8	7

No ponto horizontal se representa a variável X e no ponto vertical se representa a variável Y, como podemos ver na figura a seguir:

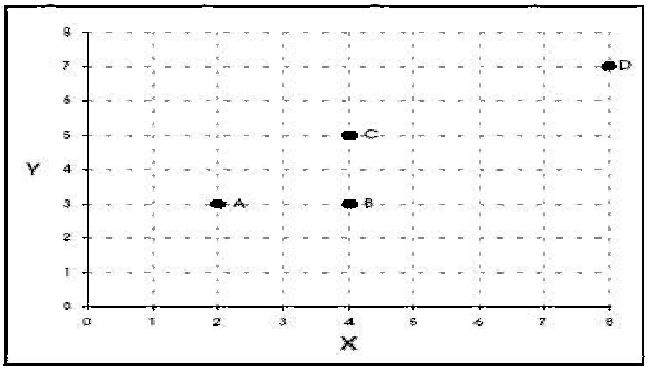

Figura 1: **Modelo do diagrama de dispersão**

FONTE: Modelo baseado em (PEÑA; ROMO, 1997, p. 411).

As medidas de assimetria são aquelas medidas (índice percentílico de assimetria, primeiro coeficiente de assimetria de Pearson, segundo coeficiente de assimetria de Pearson, índice momento de assimetria, dentre outros) que possibilitam analisar uma distribuição de acordo com as relações entre suas medidas de moda, média e mediana.

As medidas de curtose indicam o grau de achatamento da distribuição ou o quanto uma curva de freqüência será achatada em relação a uma curva normal de referência. Para o cálculo do grau de curtose de uma distribuição utiliza-se o coeficiente de curtose (ou coeficiente percentílico de curtose). Quanto a curtose a distribuição pode ser: mesocúrtica, platicúrtica e leptocúrtica.

5. Números-índices

Os números-índices são medidas estatísticas utilizadas para comparar grupos de variáveis relacionadas entre si ou evoluções de variáveis através do tempo para obter um quadro simples e resumido de mudanças significativas em áreas definidas previamente. Mediante o emprego desta técnica é possível estabelecer comparações entre:

a) Variações ocorridas ao longo do tempo;

b) Diferença entre lugares;

c) Diferenças entre categorias semelhantes, como produtos, pessoas e organizações (PEÑA; ROMO, 1997).

Cada número índice de uma série costuma vir descrito em termos percentuais. Os índices medem variações ao longo do tempo. Na Biblioteconomia e Documentação os números índices são empregados para controlar as variações relativas ao preço, às quantidades e os valores de livros, serviços ou produtos de informação.

Para os valores de uma determinada magnitude ou índice projetado ano a ano, sua taxa de variação relativa se obtém dividindo a variação absoluta entre dois anos pela quantidade associada ao primeiro deles (ESCUDER VALES; MURGUI ESQUERDO, 1995). A quantidade total de recursos financeiros gastos a cada ano, em relação ao ano base, varia de um ano para outro devido às alterações do número de unidades comparadas das diferentes variáveis e igualmente devido às mudanças nos preços unitários dos mesmos.

Uma aplicação prática dos números índice relativa ao preço é identificada juntamente com o fator tempo. Exemplo:

a) O preço de determinado jornal em 1979 foi R$ 1,20 e em 1980 subiu para R$ 1,38;

b) para o ano considerado base corresponderá sempre um índice igual a 100. Os demais apresentarão valores que flutuam em torno de 100;

c) p(79,80) = p1980 = 1,38 / 1,20 = 1,15 ou 115%, onde o resultado indica que em 1980 houve um aumento de 15% no preço do jornal em relação a 1979.

Outra possibilidade é fazer uma comparação de bens em forma de relação quantitativa, também em variação de anos diferentes.

Exemplo:

a) Um Centro de Documentação descartou 45 toneladas de documentos em 1998 e 68 toneladas em 1999;
b) A quantidade relativa será estipulada em torno do ano 1998;
   q (98,99) = q1999 = 68 / 45 = 1,51 ou 151%;
c) No ano de 1999 o Centro de Documentação descartou em média 51% a mais em relação ao ano de 1998.
d) O índice de valor determina o preço relativo de ano em ano, como nos demais índices, e está representado por: $v(o,t) = \_vt\_ = pt \cdot qt = \frac{po,t \cdot q\,o,t}{vo \quad po \cdot qo.}$

Exemplo:

a) Uma Biblioteca comprou em 1998 a quantidade de 1000 unidades de artigos de revistas a um preço unitário de R$ 500,00 de uma empresa de publicações científicas. No ano seguinte (1999) comprou 2000 unidades de artigos de revistas a um preço unitário de R$ 600,00 da mesma empresa;
b) O valor relativo em 1999 foi;
c) v (98, 99) = 600 * 2000 =120000;
d) por outro lado se calcula 500 * 1000 = 50000;
e) 120000 / 50000 = 2,4 ou 240%;
f) Em 1999 o índice de valor referente a compra de artigos de revistas foi 240% superior ao ano de 1998. Contudo, alguns índices agregados não satisfazem essa propriedade (ESCUDER VALES; MURGUI ESQUERDO, 1995).

Outro emprego do índice é o ponderado ou agregado, que não contempla o mesmo aspecto dos números índices, no entanto mantêm a mesma filosofia estatística. Nos índices ponderados ou agregados (P (P1Q/Q0) x 100) as

fórmulas são utilizadas para interpretar as variações dos preços e as quantidades dos bens. A ponderação se constitui pelos métodos baseados na participação de cada bem no valor de transação total.

Os métodos de aplicação destes índices são:

a) Método Laspeyres: utiliza as quantidades consumidas durante o período base, é o mais usado, devido a requerer medidas de quantidades unicamente de um período. Como cada número índice depende dos mesmos preços e quantidade base, a administração pode comparar o índice de um período diretamente com o índice de outro. Uma vantagem deste método é a comparabilidade de um índice com outro. O uso da mesma quantidade de período base permite fazer comparações de maneira direta. Outra vantagem é que muitas medidas de quantidade de uso comum não são tabuladas a cada ano. A principal desvantagem é que não são tomadas em conta as mudanças dos padrões de consumo (ESCUDER VALES; MURGUI ESQUERDO, 1995). Exemplo: dado o ano de 1999 como base para compra da base de dados EconLit, portanto os aumentos não serão calculados pela forma impressa (editada a muito mais tempo) e sim pela primeira aquisição da biblioteca desta base de dados;

b) Método de Paasche: é um processo parecido ao seguido para encontrar o índice de Laspeyres. A diferença consiste em que os pesos utilizados no método Paasche são as medidas de quantidade correspondentes ao período atual. É particularmente útil porque combina os efeitos das mudanças de preço e dos padrões de consumo, pelo que é um melhor indicador das mudanças gerais da economia do que o método Laspeyres. Uma das principais desvantagens é a necessidade de somar medidas de quantidade para cada período examinado. Cada valor de um índice de preços Paasche é o resultado tanto de mudanças no preço como na quantidade consumida correspondente ao período base. Como as medidas de quantidade utilizadas por um período de índice, em geral, são diferentes das medidas de quantidade de outro período de índice, resulta impossível atribuir a diferença entre os dois índices somente a mudanças de preço. Em conseqüência, é difícil comparar índices de diferentes períodos com o método Paasche (CRESPO, 2002).

6. A aplicação na bibliometria e na cienciometria

A estatística está presente em quase todas as ciências, seja de uma forma progenitora ou simplesmente como uma disciplina a mais. Dentro da Biblioteconomia e Documentação sua representação se dá, sobretudo, na Bibliometria, que é a disciplina responsável de tratar e medir a atividade científica e social através do estudo e análise da literatura consultada em qualquer tipo de suporte (SANZ-CASADO; MARTÍN MORENO, 1998).

A Bibliometria é o vínculo entre os indicadores de Ciência, Tecnologia e Inovação (que agregam os indicadores de input e os indicadores de produção científica) com a Cienciometria. Portanto, a evolução lógica da Bibliometria é a Cienciometria.

Por outro lado, a Cienciometria pretende identificar as leis e as regularidades que regem a atividade científica (CALLON; COURTIAL; PENAN, 1995). Sua aplicação é a avaliação da ciência pela ciência e sua divisão está efetuada em quatro pontos-chaves: atividade científica, crescimento exponencial, comunidade científica e os colégios invisíveis. Portanto, quando se fala de agentes de política científica e gestão da ciência, indiretamente, está aplicando-se técnicas de Cienciometria em seus relatórios.

A contribuição da Bibliometria e da Cienciometria à Ciência da Informação consiste no tratamento matemático ou estatístico da informação e da documentação, contribuindo com sua base teórica, tornando possível que se enlacem os resultados entre a ciência e sua respectiva literatura, se aceita esta relação, é possível traçar uma relação concreta entre o conhecimento e a informação registrada (PRICE, 1964).

A medição do esforço e repercussão da atividade científica na atualidade se estabelece a partir destes dois campos de estudo métrico (a Bibliometria e a Cienciometria). Os indicadores que se constroem a partir destas técnicas quantificam a quantidade de documentos publicados por país, instituição, grupos de investigação ou indivíduos, bem como o progresso científico de uma área. No entanto, as medidas mais comuns são as baseadas nas publicações e nas citações (LICEA DE ARENAS; SANTILLÁN-RIVERO, 2002).

Estas medidas não se resumem a avaliação da produção científica (publicações) e às análises de citações (as citações ou os consumos de informação), principalmente pelo crescimento científico nas últimas décadas, bem como sua recompilação em bases de dados bibliográficas automatizadas, potenciando o uso dos estudos métricos e a geração de indicadores para medir os resultados da atividade científica.

Dentro dos indicadores, é importante destacar aqueles que fundamentalmente cumprem com a finalidade de apontar os resultados imediatos e os efeitos do impacto do esforço destinado à C&T+I, constituindo-se em indicadores-produto e, em algumas situações, em medidas de impacto das políticas científicas (JANNUZZI, 2002). Assim, os indicadores bibliométricos são indicadores de eficácia quando se faz referência a resultados mais imediatos das políticas na produção de artigos em C&T ou na vigilância tecnológica.

Por outro lado, dentro dos indicadores bibliométricos, existem também os indicadores de impacto (ou indicadores de efetividade social), que são utilizados para o fomento das atividades de C&T, como o fator impacto de publicações e outras medidas como a taxa de inovação tecnológica, o balanço de inputs e outputs tecnológicos, o grau de apropriação de tecnologia, e o desenvolvimento de novos conhecimentos.

Seguindo a vertente dos indicadores bibliométricos é vital ressaltar outros dois indicadores utilizados com muita freqüência:

a) input e output: o primeiro é um indicador do volume de investimento em pesquisa científico-tecnológica (RUIVO, 1994), o segundo é um indicador de saída – utilizado para avaliar o retorno dos investimentos, e diz respeito a indicação da quantidade de livros ou artigos publicados, quantidade de citações do artigo, entre outros (WHITE; McCAIN, 1989). Estes indicadores são mais visíveis nos relatórios de investimentos públicos e privados em pesquisa de C&T+I, representados através do número de institutos, universidades, grupos de investigação e literatura branca e cinzenta. Entretanto, podem ser aplicados nas unidades de informação para verificar o grupo de usuários potenciais e a quantidade de documentos retirados (input e output simplificado);

b) processo: que são medidas destinadas a calcular os recursos em C&T+I, representados, por exemplo, pela taxa de titulação de doutores, as matrículas em cursos de Pós-graduação, a realização de congressos e exposições científicas. Sua função é interpretar os indicadores através da gestão, reprodução, disseminação e o aperfeiçoamento da política científico-tecnológica (MUGNAINI; JANNUZZI; QUONIAM, 2004). Para as unidades pode ser representado pelos níveis de clientes que ela possui.

7. Considerações finais

Todos os tipos de indicadores mencionados até agora podem ser utilizados para: (i) analisar as políticas científicas, que reafirmam a necessidade de propor objetivos e metas a atingirem; (ii) analisar a produtividade científica, justificando os investimentos financeiros que os pesquisadores recebem; (iii) avaliar os índices congruentes (indicadores), que representam a visibilidade científica.

Os indicadores são uteis para facilitar a avaliação do retorno dos recursos investidos em compra de materiais e de que forma estão sendo utilizados tais materiais. Parece sem lógica, porém dentro de uma análise métrica de consumo de informação é possível saber se alguns títulos estão sendo indexados de forma correta ou não, pela sua freqüência de saída, uso e pelos usuários que a estão utilizando (quais áreas do conhecimento).

Especificamente, podemos utilizar a base teórica de todos os indicadores bibliométricos, para uma aplicação mais direcionada, tal como indica Gorbea Portal (2005), em suas representações dos estudos métricos, através de alguns índices, como:

a) O Modelo de Price, que permite uma análise que identifica a elite de autores mais importantes de uma determinada área, para o que se utiliza a soma da elite de autores que publicam 50% dos trabalhos e se eleva à população total de autores; $E = \sqrt{N}$. Para as unidades de informação pode ser aplicado quanto a freqüência de títulos retirados do acervo, quanto ao nome dos autores e verificar de que forma os autores são representados, como um ranking, aonde teremos os autores mais consultados;

b) O Modelo matemático de Bradford que determina o núcleo das revistas mais produtivas por temas e cuja representação se consegue a partir da quantidade de títulos por zonas, multiplicado pelo fator de proporcionalidade de títulos entre as zonas. Muitas vezes se calcula a soma de revistas por uma casta do maior ao menor, em três dimensões. O cálculo é uma divisão do total por 3, formando uma zona de grande percentagem representativa (1ª escala), uma outra zona de média percentagem representativa (2ª escala) e uma terceira zona de pequena percentagem representativo (3ª escala); p: p1: p2: 1: n: n2. Esta regra é utilizada para verificar quais títulos de revistas são mais importantes para serem assinados por uma biblioteca. Com o advento do Portal de Periódicos CAPES esta análise quase não é realizado em centros públicos, entretanto para unidades particulares é uma solução viável;

c) O Modelo de Lotka veio para atender à capacidade estatística da produção por autoria, estabelecendo os fundamentos da lei do quadrado inverso (LOTKA, 1926), afirmando que o número de autores que fazem “n” contribuições num determinado campo científico é aproximadamente $1/n^2$, enquanto os que fazem somente uma contribuição são mais ou menos 80%. Parâmetros, utilizados posteriormente para medir o grau de importância dos autores em suas disciplinas e suas correlações com outros autores, em grau de proximidade. Pode ser utilizado nas unidades de informação para dizer quem são os referencias teóricos de uma área, disciplina ou ciência;

d) O Modelo de Zipf estabeleceu uma comparação das palavras-chave com as citações, onde observou que o uso das palavras em qualquer língua está claramente definido por valores constantes. Assim, definiu que o número Y de vezes que aparece uma palavra é inversamente proporcional ao seu ranking X, isto é, Y=a X/. Para uma unidade de informação é importante para verificar como se comporta a área, por exemplo, dentro da Web of Science ou até no Google Acadêmico. Pode parecer desnecessário, mas existem muitas áreas em consolidação, todavia, por este motivo é fundamental fazer de vez em quando uma análise deste tipo. Serve também para o processo de indexação de terminologias.

Referências

BABBIE, Earl. *Métodos de Pesquisas de Survey*. Belo Horizonte: Editora UFMG, 2001

CALLON, Michel; COURTIAL, Jean-Pierre ; PENAN, Hervé. *Cienciometría: el estudio cuantitativo de la actividad científica – de la Bibliometría a la vigilancia tecnológica*. Gijón: Ediciones TREA, 1995. 110 p.

CRESPO, Antonio Argot. *Estatística fácil*. São Paulo: Saraiva, 2002. 224 p.

ESCUDER VALLES, Roberto; MURGUI IZQUIERDO, J. *Santiago. Introducción a la estadística para las Ciencias Sociales*. Aravaca: McGraw-Hill, 1995. 428 p.

ETHERINGTON, Sue. *Como criar tabelas e gráficos*. São Paulo: PubliFolha, 2000. 72 p.

FERNÁNDEZ PALACÍN, Fernando et al. *Estadística descriptiva y probabilidad (teoría y problemas)*. Cádiz: Universidad de Cádiz, 2000. 257 p.

GORBEA PORTAL, Salvador. *Modelo teórico para el estudio métrico de la información documental*. Madrid: Ediciones TREA, 2005. 176 p.

JANNUZZI, Paulo Martino. Considerações sobre o uso, mau uso e abuso de indicadores sociais na avaliação de políticas públicas municipais. *Revista de Administração Pública*, Rio de Janeiro, v. 36, n. 1, p. 51-72, 2002.

LEVIN, Jack; LEVIN, William C. *Fundamentos de estadística en la investigación social*. México: Oxford University, 1997. 305 p.

LICEA DE ARENAS, Judith; SANTILLÁS-RIVERO, Emma Georgina. Bibliometría ¿para qué?. Biblioteca Universitaria, *Nueva Época*, v. 5, n. 1, p. 3-10. 2002. Disponível em: <http://www.dgbiblio.unam.mx/servicios/dgb/publicdgb/bole/fulltext/volV12002/pgs-03-10.pdf>. Último acesso em 20/05/2010.

LOTKA, Alfred J. The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, Washington, v. 16, n. 12, p. 317-323, 1926.

MARÍN FERNÁNDEZ, Josefa. *Estadística aplicada a las Ciencias de la Documentación.* Murcia: Diego Marin, 2000. 487 p.

MUGNAINI, Rogério; JANNUZZI, Paulo Martino; QUONIAM, Luc Marie. Indicadores bibliométricos da produção científica brasileira: uma análise a partir da base Pascal. *Ciência da Informação*, Brasília, v. 33, n. 2, p. 123-131, 2004.

PEÑA, Daniel; ROMO, Juan. *Introducción a la estadística para las Ciencias Sociales*. Aravaca: McGraw-Hill, 1997. 428 p.

PRICE, Derek J. de Solla. *Little science, big science*. New York: Columbia University, 1964. 119 p.

RUIVO, Beatriz. Phases or paradigms of science policy? *Science and Public Policy*, Guildford, v. 21, n. 3, p. 157-164, 1994.

SANCHEZ CORONA, Octaviano. *Probabilidad y Estadística*. México: McGraw Hill, 2005. 303 p.

SANZ-CASADO, Elías; MARTÍN MORENO, Carmen. Aplicación de técnicas bibliométricas a la gestión bibliotecaria. *Investigación Bibliotecológica*, México D.F., v. 12, n. 24, p. 24-40, 1998.

WHITE, Howard D.; McCAIN, Katherine W. Bibliometrics. *Annual Review of Information Science and Technology (ARIST)*, Medford, v. 24, p. 119-186, 1989.

Dados dos autores

Adilson Luiz Pinto

Possui graduação em Biblioteconomia pela Pontifícia Universidade Católica de Campinas (2000), mestrado em Ciência da Informação pela Pontifícia Universidade Católica de Campinas (2004) e doutorado em Documentación pela Universidad Carlos III de Madrid (2007). Tem experiência na área de Ciência da Informação, com ênfase em Representação e Organização da Informação, atuando principalmente nos seguintes temas: Estudos Métricos da Informação (bibliometria, cienciometria, informetria e webometria) e da Documentação (Arquivometria), base de

dados, recuperação de informação, fontes de informação voltados a mineração de dados para os estudos métricos e análise de redes sociais.

adilson@cin.ufsc.br

Alexandre Oliveira de Meira Gusmão

Possui graduação em Curso de Biblioteconomia pela Universidade Federal de Pernambuco(1995) , mestradoem Ciência da Informação pela Universidade Federal da Paraíba(2001) , doutorado em Documentación pela Universidad Carlos III de Madrid(2012) , curso-tecnico-profissionalizantepela Escola Técnica Estadual Professor Agamemnom Magalhães(1989) , curso-tecnico-profissionalizantepelo Centro Federal de Educação Tecnológica de Pernambuco(1992) e ensino-fundamental-primeiro-graupela Escola Amaury de Medeiros(1985) . Atualmente é Professor universitário da Universidade Federal de Mato Grosso. Tem experiência na área de Ciência da Informação. Atuando principalmente nos seguintes temas: Gestão da informação, Pecuaria de corte.

aomgusmao@hotmail.com

André de Souza Pena

Possui graduação em Biblioteconomia pela Universidade Federal de Minas Gerais (2002) e mestrado em Ciências da Informação pela Universidade Federal de Minas Gerais (2007). Tem experiência na área de Biblioteconomia e Ciência da Informação, com ênfase em Informação e Trabalho, atuando principalmente nos seguintes temas: profissional da informação, mercado de trabalho, bibliotecários, formação profissional. Atualmente é estudante de doutorando na Universidade Federal de Minas.

andresouzapena@gmail.com

Marcelo Moreira Ferreira da Silva

Atualmente é mestrando no Curso de Mestrado em Economia Aplicada da Universidade Federal de Juiz de Fora - Minas Gerais. Possui graduação em Estatística pela Universidade Federal de Minas Gerais (2010). Atuou como estatístico em cargo comissionado do Governo de Minas Gerais na Secretária de Estado de Desenvolvimento Social de Minas Gerais auxiliando na construção de indicadores sociais. Atuou como auxiliar de pesquisa da Fundação João Pinheiro para construção de indicadores no setor de Contas Regionais e municipais. Tem experiência na área de Probabilidade e Estatística, com ênfase em Probabilidade e Estatística Aplicadas, atuando principalmente nos seguintes temas: rna, fisher, mlg, estimativas e schwartz.

marcelloest@gmail.com

Recibido - Received: 2011-08-26
Aceptado - Accepted: 2012-03-30