

# Estudo comparativo teórico metodológico dos índices normalizados de citação: uma aplicação na ciência brasileira (1996-2007)

**Maria Cláudia Cabrini Grácio**

Doutora em Lógica; UNESP-Marília;  
cabrini@marilia.unesp.br

**Ely Francina Tannuri de Oliveira**

Livre Docente em Estudos Métricos; UNESP-Marília;  
etannuri@gmail.com

**Resumo:** Objetiva-se apresentar, analisar e comparar os resultados da aplicação de três procedimentos de normalização do índice de citação por documento: normalização pela média geral, pela mediana e pela média dos 10% mais produtivos da área, no universo da ciência brasileira nas 27 áreas do conhecimento, apresentadas pela Scopus. A partir dos dados do portal SCImago Journal & Country Rankings para as 27 áreas, calcularam-se Ma, Md e Ma10% para o número de citações por documento para o conjunto de países. A seguir, para cada área, calculou-se o índice normalizado do Brasil pela média geral ( $IN_{ma}$ ), pela mediana ( $IN_{md}$ ) e pela média dos 10% dos países mais produtivos ( $IN_{ma_{10\%}}$ ). Concluiu-se que os procedimentos de normalização apresentam resultados com menos vieses e mais representativos do desempenho real da área em relação aos pares, quando se comparam diferentes áreas do conhecimento. Essa é a grande contribuição dos indicadores normalizados apontada por esta pesquisa.

**Palavras-chave:** Estudo teórico-metodológico. Índices de citação. Índices normalizados de citação.

## 1 Introdução

A ciência brasileira tem crescido de forma expressiva, particularmente nos últimos 25 anos (ZANOTTO, 2002; LETA; CHAIMOVICH, 2002; LETA; GLÄNZEL; THIJS, 2006; LEITE; MUGNAINI; LETA, 2011), levando o Brasil à posição de liderança no contexto da comunidade científica latino-americana (GLÄNZEL; LETA; THIJS, 2006).

Por meio de indicadores absolutos e normalizados, os estudos cientométricos destacam-se dentre os métodos para a análise e avaliação da produção científica, contribuindo para a visualização do comportamento da ciência nas diferentes áreas do conhecimento. Os indicadores normalizados identificam, de forma comparativa, questões relativas à produção e ao impacto científico atinentes às diferentes áreas do conhecimento, na medida em que padronizam as unidades de medida (GLÄNZEL et al., 2009).

Esta pesquisa tem por objetivo geral realizar um estudo comparativo da contribuição dos índices normalizados de citação para a visualização do impacto da ciência produzida, analisando a presença brasileira nas 27 áreas do conhecimento apresentadas pelo Portal *SCImago Journal & Country Rank* no período de 1996 a 2007.

De forma mais específica, o estudo propõe-se a apresentar, analisar e comparar os resultados da aplicação de três procedimentos utilizados na normalização do índice de citação por documento, a saber: normalização pela média geral, pela mediana e pela média dos 10% mais produtivos da área, considerados critérios mais fidedignos para a comparação entre áreas. Ainda, objetiva-se avaliar a correlação entre o índice absoluto de citações por documento e os índices normalizados pelos três procedimentos, analisando se esses indicadores evidenciam a mesma tendência ou independência como descritores do desempenho científico das diferentes áreas.

Esta pesquisa se justifica por contribuir, de forma analítica, para o desenvolvimento da temática em estudo ao oferecer subsídios teóricos e metodológicos, fortalecendo sua compreensão e uso, bem como pela necessidade de estudos eminentemente brasileiros sobre o tema, contribuindo como aporte para as políticas científicas institucionais e governamentais.

## **2 Índices cientométricos normalizados de citação**

A análise da produção científica envolve um grupo de indicadores bibliométricos

absolutos e normalizados, que podem ser agrupados em indicadores de produção, citação e ligação (GREGOLIN et al., 2005). Nesta pesquisa, enfocam-se os indicadores de citação.

Os indicadores absolutos de citação permitem identificar a visibilidade e o impacto de um pesquisador ou país dentro de uma comunidade científica. Entre eles, destaca-se o número de citações por documento de um pesquisador, instituição ou país. Calculado pela média de citações, é um indicador que contextualiza o impacto de um pesquisador, instituição ou país ao relativizar as citações em função do número de documentos produzidos.

Apesar da sua importância, os indicadores absolutos de citação têm diversas limitações quando usados para comparar o desempenho entre diferentes áreas. Essa questão decorre das especificidades das áreas, uma vez que as disciplinas científicas diferem na prática de citação, tendo como consequência o fato de o número de citações recebidas por um artigo em uma dada área poder diferir do número de citações típico de outra área.

Assim, os indicadores absolutos podem levar a avaliações injustas quando se realizam análises comparativas das áreas, resultando em análises distorcidas da inserção e impacto das diferentes áreas. Autores como Persson, Glänzel e Danell (2004) apontam que o uso desses indicadores pode conduzir a conclusões equivocadas, e recomendam o uso de indicadores normalizados.

Neste contexto, Vinkler (2012) sugere a aplicação de indicadores cientométricos normalizados nas avaliações comparativas, a fim de se eliminar a influência das particularidades e especificidades das diferentes áreas científicas. Segundo o mesmo autor, processos de normalização dos indicadores de impacto têm sido amplamente aplicados em Cientometria há algum tempo, e o mesmo chama a atenção para a existência de vários tipos de índices normalizados em função de diferentes parâmetros de normalização.

Há uma tendência no sentido de propor procedimentos de normalização que possibilitem eliminar a dependência do contexto da área e, assim, oferecer critérios mais fidedignos na comparação entre disciplinas, áreas, instituições ou em nível

macro, uma vez que padronizam as unidades de medida (GLÄNZEL et al., 2009, LI et al., 2013).

Dentre os procedimentos de normalização propostos, destacam-se a normalização pela média da área ( $IN_{ma}$ ) e pela mediana da área ( $IN_{md}$ ) (MOED, 2009, LI et al., 2013). Um terceiro procedimento de normalização a ser apresentado neste estudo é pela média dos 10% mais produtivos ( $IN_{ma_{10\%}}$ ), uma adaptação daquele presente em Moed (2010), em que o autor se refere ao conjunto de 10% dos artigos mais citados, considerando-se que a base utilizada na pesquisa não apresenta o número de citações individualmente nos artigos, mas a média de citações por artigo por país. Optou-se, assim, pelos 10% dos mais produtivos em lugar dos 10% com maior média de citação, pois pode ocorrer baixa produtividade e altos valores de citação, decorrentes, muitas vezes, de colaboração com países altamente produtivos.

Para estudos comparativos, um indicador normalizado pode ser definido como o quociente entre o valor do indicador obtido para o indivíduo (seja uma instituição ou país) dividido pelo parâmetro de normalização, seja a média, a mediana ou a média dos 10% mais produtivos (MOED, 2009, LI et al., 2013). Como resultado, esse indicador padroniza o comportamento de um indivíduo, de forma a situá-lo em relação à tendência global (parâmetro utilizado) observada na área. Nesta pesquisa, os indivíduos se referem a países.

O indicador normalizado é calculado pela fórmula:

$$IN_j = \frac{I_j}{PN_g}$$

em que:

$IN_j$  = índice normalizado para o país  $j$ ;  $I_j$  = valor do indicador absoluto para o país  $j$ ;  
 $PN_g$  = parâmetro de normalização - média ( $Ma$ ), mediana ( $Md$ ) ou média dos 10% mais produtivos do indicador da área em estudo ( $Ma_{10\%}$ ).

O valor 1 obtido para o índice normalizado (IN) indica que o desempenho do país (nesta pesquisa, o Brasil) é igual ao comportamento de referência ( $Ma$ ,  $Md$  ou  $Ma_{10\%}$ ) do grupo (nesta pesquisa, os países) de produtores na área. Valores abaixo

de 1 significam que o país está abaixo do que é esperado para o grupo, na área em estudo; valores acima de 1 apontam que o país apresenta um desempenho acima do comportamento referente ( $Ma$ ,  $Md$  ou  $Ma_{10\%}$ ) do grupo.

Instituições, tais como a *Tampere University of Technology* (2011), sugerem que os valores dos índices normalizados pela média podem ser categorizados em três estratos de impacto: para valores menores que 0,80, baixo impacto; valores entre 0,80 e 1,20, com médio impacto; e os que estão acima de 1,20, alto impacto. No presente estudo, considera-se que esses parâmetros podem ser extensivos aos outros procedimentos de normalização.

### 3 Procedimentos metodológicos

Em dezembro de 2013, no portal *SCImago Journal & Country Rankings*, em *Country Rankings*, para cada uma das 27 áreas, fez-se o *download* dos dados em formato Excel referentes ao período 1996-2007, selecionando-se os dados relativos ao total de documentos produzidos e o número de citações por documento para o conjunto de países produtores.

Para as 27 áreas, calcularam-se  $Ma$ ,  $Md$  ou  $Ma_{10\%}$  para o número de citações por documento para o conjunto de países. A seguir, para cada área, calculou-se o índice normalizado do Brasil pela média geral ( $IN_{ma}$ ), pela mediana ( $IN_{md}$ ) e pela média dos 10% dos países mais produtivos ( $IN_{ma_{10\%}}$ ), apresentados na Tabela 1. Para a Tabela 1, calcularam-se média, desvio padrão, mínimo, máximo e coeficiente de variação (C.V.) em relação ao conjunto de 27 áreas para cada um dos índices, a partir dos quais se descrevem as tendências observadas, assim como suas similaridades e diferenças.

Para complementar a visualização das tendências dos índices absoluto e normalizados de citação, foram construídos dois gráficos do tipo Diagrama de Caixa, também denominado *Box Plot*, no software SPSS, que destacam os valores mínimo, 1º, 2º (ou mediana) e 3º Quartis, e máximo, assim como os valores discrepantes (*outliers*) de cada índice de citação. Cada caixa, nessa representação

gráfica, é composta por dois retângulos, em que o primeiro representa a distância entre a Mediana (2º Quartil) e o 1º Quartil, e o segundo, a distância entre o 3º Quartil e a Mediana. A partir dos 1º e 3º Quartis são desenhadas linhas verticais até os valores mínimo e máximo para os índices de citação absoluto e normalizados. Os valores discrepantes, quando existentes, aparecem distantes das caixas.

A seguir, calcularam-se as correlações de Pearson entre o índice absoluto de citações por documento e os índices normalizados pelos três procedimentos, realizando as respectivas análises. A correlação positiva entre dois índices indica que valores pequenos (ou grandes) observados para um dos índices tendem a estar relacionados a valores pequenos (ou grandes) no outro índice. Por outro lado, uma correlação negativa entre dois índices indica que valores pequenos (ou grandes) para um índice tendem a estar relacionados a valores grandes (ou pequenos) para o outro índice. Ainda, quanto mais próximo o coeficiente de correlação ( $r$ ) for do valor 1 (ou  $-1$ ), mais intensa é a correlação entre os dois índices; se próximo a zero, evidencia-se a ausência de correlação (BARBETTA, 1999).

A distribuição dos índices de citação, apresentados dois a dois, e respectivos valores para correlações foram representados em seis gráficos de dispersão, inserindo-se a respectiva linha de tendência em cada um deles, utilizando-se o software *SPSS*.

#### 4 Apresentação e análise de dados

Apresenta-se a Tabela 1 com os valores dos índices de citações por documento, absoluto e normalizados, para as 27 áreas, ordenados em função do  $IN_{ma}$ , com as áreas de Enfermagem e de Economia, Econometria e Finanças apresentando os maiores índices de  $IN_{ma}$ . Em relação à área de Enfermagem, os valores (8,8; 6,2; 19,8) representam a média, a mediana e a média dos 10% mais produtivos de citação do conjunto de países produtores na área. Assim, o valor de 1,8 para  $IN_{ma}$  indica que a produção científica brasileira em Enfermagem apresenta impacto superior àquele do grupo de países na área, mais especificamente, igual a 1,8 vezes o impacto

médio do grupo.

Em relação à distribuição dos índices absolutos de citações por documento, destaca-se a área Multidisciplinar, com valor próximo ao dobro da área de Ciências Ambientais, que tem o segundo maior valor observado. Além dessa, há 11 áreas com índice acima da média calculada (11,3). As outras 15 áreas apresentam citações por documento abaixo da média, e a área de Artes e Humanidades com o menor índice para esse indicador (2,9). Destaca-se, ainda, a grande dispersão deste índice, identificado pelo alto valor de coeficiente de variação (C.V.) (acima de 30%), apontando a não representatividade da média para a descrição do mesmo.

A partir dos valores calculados para  $IN_{ma}$ , somente sete áreas apresentam valor menor que um, indicando que o impacto das citações destas áreas está abaixo da média dos países produtores. Por outro lado, 15 áreas apresentam valor para esse índice acima de 1, significando que para essas áreas o país apresenta um desempenho acima do comportamento médio do grupo produtor.

**Tabela 1** - Índices absoluto e normalizados de citação por documento para as 27 áreas

Área (média; mediana; média 10%)	Cit/doc	$IN_{Ma}$	$IN_{Md}$	$IN_{Ma10\%}$
Enfermagem (8,8; 6,2; 19,8)	8,80	15,80	1,80	2,50
Economia, Econometria e Finanças (7,16; 6,72; 12,23)	7,16	12,60	1,80	1,90
Odontologia (11,28; 9,53; 15,89)	11,28	15,20	1,40	1,60
Negócio, Gestão e Contabilidade (6,39; 4,93; 11,45)	6,39	9,20	1,40	1,90
Ciências Ambientais (12,93; 11,99; 16,71)	12,93	16,40	1,30	1,40
Psicologia (8,82; 7,45; 14,43)	8,82	11,50	1,30	1,60
Ciências da Decisão (8,17; 7,56; 11,56)	8,17	10,20	1,30	1,40
Matemática (6,06; 5,29; 9,21)	6,06	7,50	1,30	1,40
Farmacologia, Toxic. e Farmácia (11,00; 10,83; 15,53)	11,00	12,60	1,20	1,20
Química (11,76; 10,12; 16,73)	11,76	13,10	1,10	1,30
Ciências da Saúde (9,70; 8,69; 16,58)	9,70	10,80	1,10	1,30
Física e Astronomia (9,28; 8,17; 14,58)	9,28	10,50	1,10	1,30
Engenharia Química (9,41; 7,69; 11,23)	9,41	10,00	1,10	1,30
Ciências dos Materiais (7,68; 7,48; 10,45)	7,68	8,10	1,10	1,10



Área (média; mediana; média 10%)	Cit/doc	IN <sub>Ma</sub>	IN <sub>Md</sub>	IN <sub>Ma10%</sub>
Ciências Sociais (5,11; 4,56; 7,78)	5,11	5,60	1,10	1,20
Imunologia e Microbiologia (15,31; 15,15;22,30)	11,33	15,00	1,00	1,00
Bioquímica, Genética e Biol. Molec.(14,64; 13,58; 22,88)	14,64	14,30	1,00	1,10
Ciências Planetárias e da Terra (12,42; 10,92; 15,86)	12,42	12,90	1,00	1,20
Engenharia (6,22; 4,99; 8,03)	6,22	6,40	1,00	1,30
Artes e Humanidades (2,92; 1,91; 4,29)	2,92	2,90	1,00	1,50
Neurociência (16,10; 15,00; 25,96)	16,10	14,10	0,90	0,90
Medicina (11,74; 11,05; 15,26)	11,74	10,00	0,90	0,90
Agricultura e Ciências Biológicas (11,05; 10,43; 14,00)	11,05	9,60	0,90	0,90
Ciência da Computação (7,37; 5,83; 10,32)	7,37	6,80	0,90	1,20
Energia (6,96; 6,69; 9,62)	6,96	6,50	0,90	1,00
Veterinária (7,06; 6,83; 8,22)	7,06	6,10	0,90	0,90
Multidisciplinar (39,30; 30,72; 76,88)	39,30	31,60	0,80	1,00
Média	10,39	11,30	1,10	1,30
Desvio padrão	6,51	5,30	0,20	0,40
Coefficiente de variação (C.V.)	62,6%	46,6%	22,0%	27,6%
Mínimo	2,92	2,90	0,80	0,90
Máximo	39,30	31,60	1,80	2,50

Fonte: elaborada pelos autores

A área Multidisciplinar, que se destacou em relação ao valor absoluto de citações por documento, apresenta o menor valor para  $IN_{ma}$ , e a área de Artes e Humanidades, com o menor valor absoluto de citações por documento, apresenta valor igual a 1. Em relação à dispersão deste índice, o valor de C.V. menor que 30% significa a representatividade da média para esse indicador.

Para os valores calculados para  $IN_{md}$ , o C.V. menor que 30% indica a representatividade da média. Somente quatro áreas tiveram valores abaixo de 1, e 20 acima deste valor, indicando que a maioria das áreas apresenta um comportamento acima do mediano, entre seus pares, com destaque para a Enfermagem, com maior índice normalizado.



Para  $IN_{ma_{10\%}}$ , observou-se que nenhuma área, no Brasil, apresentou valor acima de 1, e somente três delas apresentaram valores iguais a 1, a saber: Ciências Ambientais, Odontologia e “Economia, Econometria e Finanças”. As duas primeiras áreas, no Brasil, estão entre as 10% mais produtivas e têm impacto igual ao desempenho médio, e a última, apesar de não pertencer ao grupo dos 10% mais produtivos na área, apresenta impacto semelhante ao observado para este grupo. Por outro lado, nas áreas Multidisciplinar, Neurociências e Bioquímica, Genética e Biologia Molecular, o Brasil apresenta-se entre os países (10%) mais produtivos, porém  $IN_{ma_{10\%}}$  encontra-se abaixo do impacto médio dos mais produtivos. Para  $IN_{ma_{10\%}}$ , o C.V. é igual a 17,5 %, o menor valor para este parâmetro.

Comparando-se os índices absolutos e os normalizados, observa-se que, no Brasil, a Enfermagem é a área com maior  $IN_{ma}$  e  $IN_{md}$ , embora em valores absolutos apresente o terceiro maior valor. O índice absoluto de citação por documento da área Economia, Econometria e Finanças apresenta o 10º maior valor (12,6); mas os índices  $IN_{ma}$ ,  $IN_{ma_{10\%}}$  e  $IN_{md}$  apresentam-se com o maior valor em relação aos dois primeiros e segundo maior valor em relação ao terceiro índice, uma vez que, nesta área, Cit/doc está acima do comportamento característico (medido pela média, igual a 7,6; pela mediana, igual a 6,72; e pela média dos 10% mais produtivos, igual a 12,23) do conjunto de países produtores.

As áreas de Odontologia e Ciências Ambientais apresentam o maior valor para  $IN_{ma_{10\%}}$  e valores acima de 1 e semelhantes para os outros dois índices normalizados. A área de Artes e Humanidades apresenta o menor valor absoluto de citações por documento, mas com valor acima de 1, para o  $IN_{md}$ , e igual a 1, para  $IN_{ma}$ . Assim, algumas áreas apresentam altos índices de impacto, pelos índices normalizados, embora em valores absolutos não apresentem um destacável impacto por artigo, quando comparados às outras áreas.

Por outro lado, a área Multidisciplinar apresenta o maior valor absoluto de citações por documento (Cit/doc). Todavia, quando observados seus  $IN_{ma}$  e  $IN_{ma_{10\%}}$ , apresenta os menores valores e abaixo de 1, uma vez que o valor de

Cit/doc (igual a 31,6) brasileiro, nesta área, está abaixo do comportamento característico (medido pela média, igual a 39,3; e pela média dos 10% mais produtivos, igual a 76,88) do conjunto de países produtores, e seu  $IN_{md}$  apresenta valor igual a 1.

Considerando as situações apresentadas, destaca-se que os índices normalizados possibilitam eliminar distorções ou *viéses* em análises comparativas entre áreas, decorrentes das particularidades e especificidades de cada área.

Comparando os três índices normalizados, observa-se que, para todas as 27 áreas, o valor da mediana (Md) é menor que o valor da média (Ma) e este é menor que o valor da média dos 10% mais produtivos ( $Ma_{10\%}$ ), isto é,  $Md < Ma < Ma_{10\%}$ , dado o caráter assimétrico das distribuições de impacto (frequência de citação). Consequentemente, considerando o conceito de índice normalizado, a relação entre eles se invertem, a saber:  $IN_{md} > IN_{ma} > IN_{ma_{10\%}}$ . Assim, o índice normalizado pela mediana apresenta valores mais destacados de impacto em relação ao grupo.

De acordo com a categorização de *Tampere University of Technology* (2011), segundo o índice normalizado pela média, 9 (33%) das 27 áreas apresentam alto impacto e 18 (67%) apresentam médio impacto. Em relação ao índice normalizado pela mediana, a maioria das áreas (18) apresenta alto impacto e 9 áreas com médio impacto. Observe-se que, das 18 áreas de alto impacto pelo  $IN_{md}$ , 9 delas são de alto impacto também pela média. Desse modo, em relação a esses dois índices, nenhuma das áreas de atuação brasileira apresenta baixo impacto.

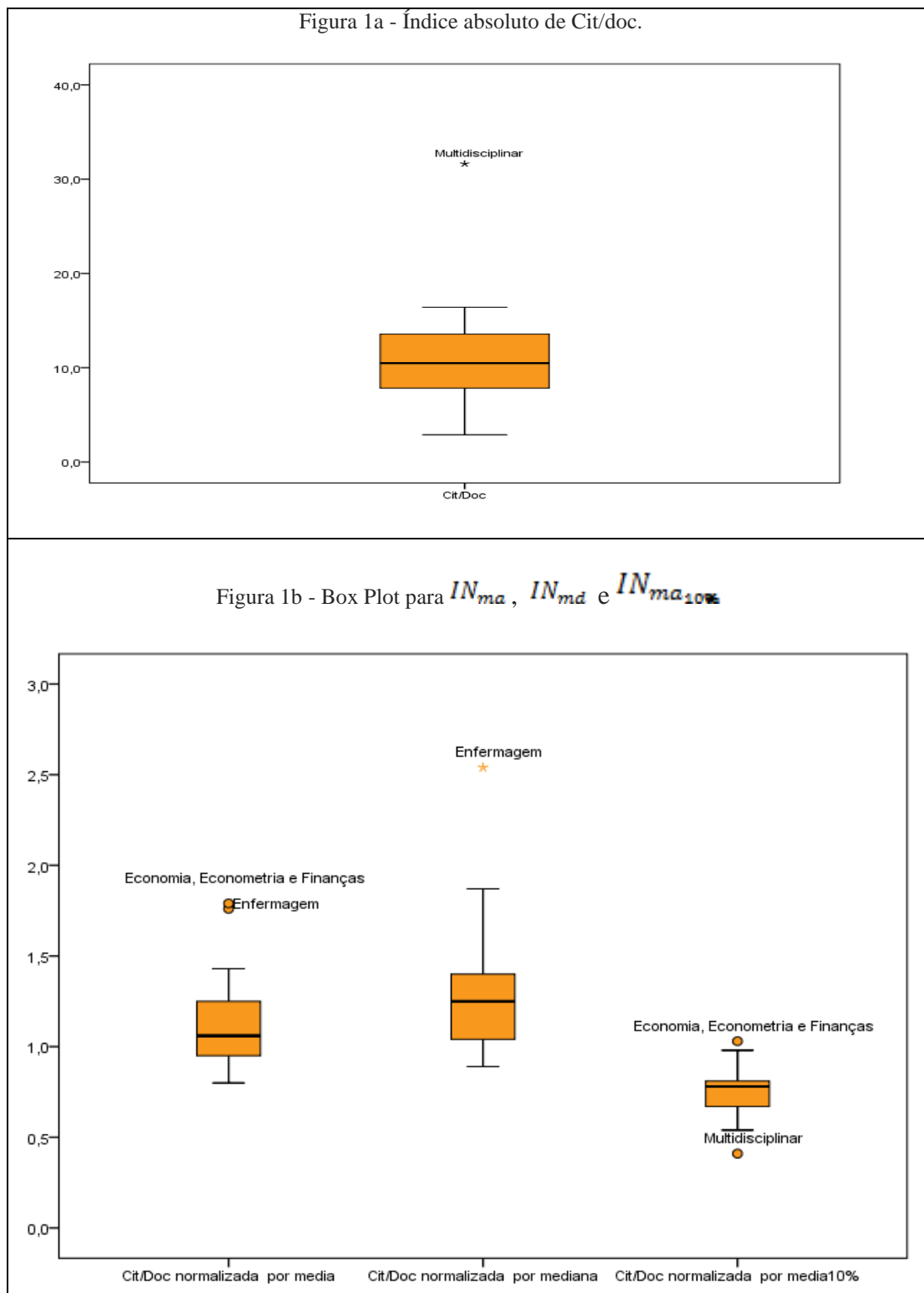
Por outro lado, quanto ao  $IN_{ma_{10\%}}$ , nenhuma área apresenta alto impacto, sendo que 14 delas apresentam médio impacto e 13, baixo impacto, em relação aos 10% dos países com maior produção nas respectivas áreas. Destaca-se que, das 27 áreas, somente 3 (Economia, Econometria e Finanças, Negócio, Gestão e Contabilidade e Ciências da Decisão) não pertencem ao conjunto dos 10% mais produtivos das suas respectivas áreas, todavia com impacto similar ao impacto médio deste grupo. As outras 24 áreas pertencem ao grupo dos mais produtivos, todavia a maioria delas apresenta a frequência de citação mais baixa do grupo ao qual pertencem. Desse modo, observa-se que os dois primeiros procedimentos de normalização ( $IN_{ma}$  e  $IN_{md}$ ) captam características distintas do último em relação ao impacto da ciência brasileira.

A Figura 1, representada pelas Figuras 1a e 1b, apresentam os Diagrama de Caixa (*Box Plot*) com a visualização das tendências dos índices absoluto e normalizados de citação, assim como os valores discrepantes (*outliers*) para cada índice.

Na Figura 1a, a área Multidisciplinar apresenta um comportamento discrepante em relação às outras 26 áreas, com valor absoluto de Cit/doc (31,6) bastante acima da tendência das áreas, que variam entre 2,9 (mínimo) e 16,4 (máximo na figura) citações por documento. Considera-se que esse comportamento pode ser decorrente da própria natureza da área, com características diversificadas, que se constitui como a confluência de áreas distintas, porém com interfaces e próximas uma das outras.

Na Figura 1b, em relação aos índices normalizados pela média, observam-se duas áreas com comportamento discrepante e acima das demais áreas, a saber: Enfermagem e Economia, Econometria e Finanças, considerando, em relação a este índice, a variação 0,8 (valor mínimo) e 1,4 (máximo).

**Figura 1 - Box plot dos índices absolutos e normalizados, pela média, mediana e média dos 10% mais produtivos de citação para as 27 áreas**



Fonte: elaborada pelos autores

Em relação ao diagrama de caixa do índice normalizado pela média dos 10% mais produtivos, observam-se dois *outliers*: Economia, Econometria e Finanças, com valor acima, e Multidisciplinar, com valor abaixo da tendência das áreas. Observe-se que as áreas de Odontologia e Ciências Ambientais, embora na Tabela 1 apresentem o mesmo valor para  $IN_{ma10\%}$ , no diagrama não aparecem como *outliers*, uma vez que seus valores, na tabela, são resultantes de arredondamentos “para cima”, igualando-se à Economia, Econometria e Finanças.

A Tabela 2 mostra os valores de correlação de Pearson entre o índice absoluto e os normalizados. O índice absoluto de Cit/doc não apresenta correlação significativa com nenhum dos índices normalizados, mostrando o comportamento independente destes.

**Tabela 2** - Correlação entre os índices absolutos e normalizados de citações por documento

Índice	$IN_{Ma}$	$IN_{Md}$	$IN_{Ma10\%}$
Cit/doc	0,05	0,05	-0,23
$IN_{Ma}$		0,92	0,7
$IN_{Md}$			0,51

Fonte: elaborada pelos autores

A maior correlação (0,92) ocorre entre  $IN_{ma}$  e  $IN_{md}$  e mostra que estes índices tendem a apresentar comportamentos semelhantes. A correlação entre  $IN_{ma10\%}$  e os outros dois índices normalizados apresenta valores significativos, porém de intensidade moderada, o que indica que o comportamento desse índice tende a ter moderada associação com os demais.

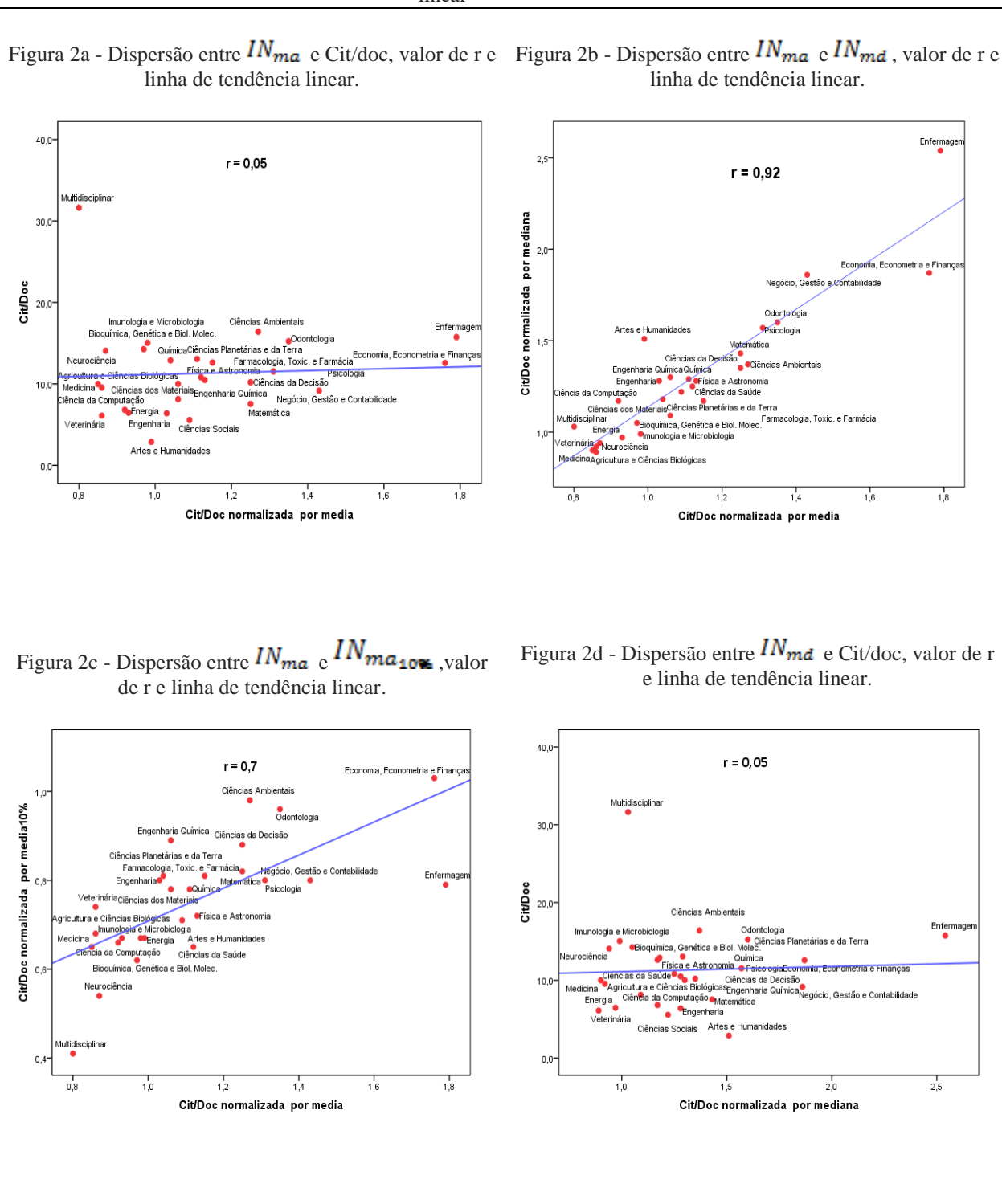
A Figura 2 apresenta seis gráficos de dispersão, com a distribuição dos valores de índices de citação, apresentados dois a dois e as respectivas linhas de regressão linear, indicando a tendência do índice de citação apresentado no eixo vertical em função do índice apresentado no eixo horizontal.

Destaca-se, inicialmente, o gráfico da Figura 2b, no qual se observa que, para 24 das 27 áreas, os valores de  $IN_{md}$  em função dos respectivos valores de  $IN_{ma}$  se apresentam aglutinados à linha de tendência linear traçada. As outras três áreas — Enfermagem, Artes e Humanidades e Economia, Econometria e Finanças — distanciam-se desta reta de regressão linear: para as duas primeiras áreas, os valores de  $IN_{md}$  apresentam valores acima da reta de tendência, enquanto, na última área, os valores de  $IN_{md}$  se encontram abaixo dela. Tais resultados sugerem a proximidade de tendência de impacto medida por esses dois índices de impacto.

As Figuras 2c e 2f apresentam a dispersão entre  $IN_{ma}$  e  $IN_{ma_{10\%}}$  e entre  $IN_{ma_{10\%}}$  e  $IN_{md}$ , com  $r$  de Pearson evidenciando uma intensidade moderada de associação entre esses índices, iguais a 0,7 e 0,51 respectivamente. A primeira (Figura 2c) apresenta aglutinação, de moderada para forte, no entorno da reta de tendência, com as áreas de Enfermagem e Multidisciplinar abaixo da mesma, e valores para Engenharia Química e Ciências Ambientais acima da reta. A Figura 2f apresenta maior dispersão dos valores em torno da reta de tendência, decorrente da moderada correlação entre os índices, com distâncias mais destacadas para as áreas de Enfermagem e Negócio, Gestão e Contabilidade, nas quais os valores obtidos para  $IN_{md}$  são maiores que aqueles esperados pela reta de tendência; e apresenta a área de Veterinária com valor de  $IN_{md}$  inferior àquele esperado para a relação de tendência linear.

Completando a análise, as Figuras 2a, 2d e 2e apresentam as distribuições de dispersão entre o índice absoluto Cit/doc e os índices normalizados de citação, e evidenciam que os valores  $IN_{ma}$ ,  $IN_{md}$  e  $IN_{ma_{10\%}}$ , apresentados no eixo horizontal, se aglutinam em torno de uma reta de tendência quase paralela a esse eixo, em consequência dos baixos valores de correlação. Nestas figuras, a área Multidisciplinar apresenta-se mais distante em relação às retas de tendência linear, com valores muito acima da tendência apontada por essas retas, constituindo um *outlier* em relação às outras áreas, como já observado na Figura 1a. Nota-se ainda, na Figura 2e, a fraca correlação negativa, representada pela reta com pequena inclinação negativa.

**Figura 2** - Gráficos de dispersão, com a distribuição dos valores de índices de citação absoluto e normalizados apresentados dois a dois, com valor de r de Pearson e respectivas linhas de tendência linear



Continua.



Conclusão.

Figura 2e - Dispersão entre  $IN_{ma_{10\%}}$  e Cit/doc, valor de r e linha de tendência linear.

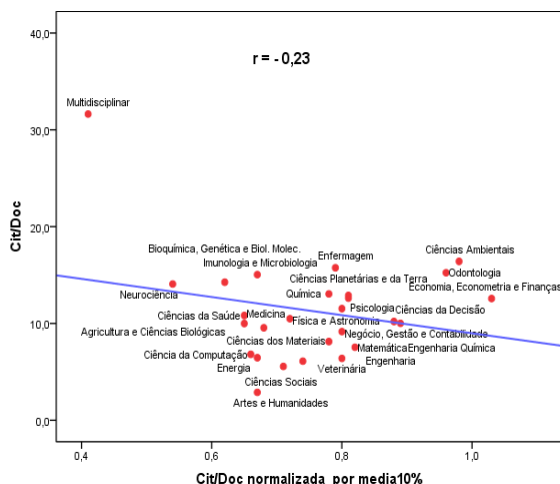
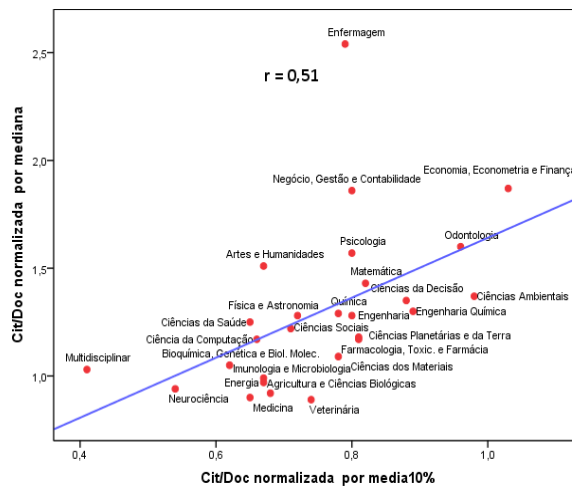


Figura 2f - Dispersão entre  $IN_{ma_{10\%}}$  e  $IN_{md}$ , valor de r e linha de tendência linear.



Fonte: elaborada pelos autores.

## 5 Conclusões

Considerando que um indicador normalizado de citação leva em conta as diferenças nas práticas de citação entre as diversas áreas do conhecimento ao relativizar o impacto de citação de um grupo atinente ao seu comportamento padrão, seja ele a média, mediana ou conjunto dos 10% mais produtivos na área em que o grupo está atuando, análises em que se utilizam esses procedimentos apresentam resultados com menos *vieses* e mais representativos do desempenho real em relação aos pares, quando se comparam diferentes áreas do conhecimento. Essa é a grande contribuição dos indicadores normalizados apontada por esta pesquisa.

A correlação entre  $IN_{md}$  e  $IN_{ma}$  mostra uma grande similaridade de comportamento destes dois índices. Por outro lado, as correlações moderadas entre  $IN_{ma_{10\%}}$  e os índices  $IN_{ma}$  e  $IN_{md}$  indicam que  $IN_{ma_{10\%}}$  pode ser considerado um índice complementar para descrever o impacto das áreas nas comunidades

científicas, ratificando a observação de Vinkler (2012) relativa ao fato de o impacto da informação científica não poder ser representado por um único índice, dada a natureza multifacetada do conhecimento científico.

À guisa de conclusão, aponta-se a contribuição desta pesquisa, de forma analítica, para o desenvolvimento da temática “Indicadores Normalizados”, ao apresentar elementos teóricos e metodológicos, fortalecendo sua compreensão e uso, propondo um estudo eminentemente brasileiro sobre o tema.

## Referências

BARBETTA, P. A. **Estatística aplicada às Ciências Sociais**. 3. ed. Florianópolis: Editora da UFSC, 1999.

GLÄNZEL, W.; LETA, J.; THIJS, B. Science in Brazil. Part 1: A macro-level comparative study. **Scientometrics**, Amsterdam, v. 67, n. 1, p. 67-86, 2006.

GLÄNZEL, W. et al. Subfield-specific normalized relative indicators and a new generation of relational charts: Methodological foundations illustrated on the assessment of institutional research performance. **Scientometrics**, Amsterdam, v. 78, n. 1, p. 165-188, 2009.

GREGOLIN, J. A. R. et al. Análise da produção científica a partir de indicadores bibliométricos. In: Francisco Romeu Landi (Org.). **Indicadores de Ciência, Tecnologia e Inovação em São Paulo 2004**. São Paulo: FAPESP, 2005. Disponível em: <<http://www.fapesp.br/2060>>. Acesso em 25 mai 2012.

LEITE, P.; MUGNAINI, R.; LETA, J. A new indicator for international visibility: exploring Brazilian scientific community. **Scientometrics**, Amsterdam, v. 88, p. 11-319, 2011.

LETA, J.; CHAIMOVICH, H. Recognition and international collaboration. **Scientometrics**, Amsterdam, v. 53, n. 3, p. 325-335, 2002.

LETA, J.; GLÄNZEL, W.; THIJS, B. Science in Brazil. Part 2: Sectoral and institutional research profiles. **Scientometrics**, Amsterdam, v. 67, n. 1, p. 87-105, 2006.

LI, Y. et al. Quantitative evaluation of alternative field normalization procedures. In: INTERNATIONAL SOCIETY OF SCIENTOMETRICS AND INFORMETRICS CONFERENCE (ISSI), 14., 2013, Viena - Austria. **Anais...** Viena: ISSI, 2013. p. 1431-1441.

MOED, H.F. New developments in the use of citation analysis in research evaluation. **Scientometrics**, Amsterdam, v. 57, p. 13 –18, 2009.

MOED, H.F. **Citation analysis in research evaluation**. Holanda: Springer, 2010.

PERSSON, O.; GLÄNZEL, W.; DANELL, R. Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. **Scientometrics**, Amsterdam, v. 60, n.3, p. 421-432, 2004.

TAMPERE UNIVERSITY OF TECHNOLOGY. **TUT Research Assessment Exercise 2011**: bibliometric report 2005-2010. 2011. Disponível em: <[http://www.tut.fi/idcprod/groups/public\\_news/@1102/@news/@p/documents/liit/021003.pdf](http://www.tut.fi/idcprod/groups/public_news/@1102/@news/@p/documents/liit/021003.pdf)>. Acesso em: 10 jan. 2014.

VINKLER, P. The case of scientometricians with the “absolute relative” impact indicator. **Journal of Informetrics**, Mumbai, v. 6, p. 254-264, 2012.

ZANOTTO, E.D. Scientific and technological development in Brazil. The widening gap. **Scientometrics**, Amsterdam, v. 55, n. 3, p. 411-419, 2002.

## **A theoretical and methodological comparative study of normalized citation indexes: an application in Brazilian science (1996-2007)**

**Abstract:** This study aims to present, analyze and compare the results of applying three normalization procedures on the index of citations per document: normalized by the overall mean, median and the mean of the 10% most productive in the area, in the universe of Brazilian science in the 27 areas of knowledge presented by Scopus. From the data obtained on SCImago Journal & Country Rankings for the 27 areas, we calculated  $Ma$ ,  $Md$  and  $Ma_{10\%}$  of the number of citations per document for the set of countries. Next, for each area, we calculated the normalized index of Brazil by the overall mean ( $IN_{ma}$ ), median ( $IN_{md}$ ) and the mean of the 10% most productive countries ( $IN_{ma_{10\%}}$ ). It was concluded that normalization procedures present results with fewer biases and are more representative of the actual performance of the area in relation to its peers when comparing different areas of knowledge. This is the major contribution of normalized indicators pointed out by this research.

**Keywords:** Theoretical-methodological study. Citation indexes. Normalized citation indexes.

Recebido: 29/07/2014

Aceito: 03/11/2014