



**UNIVERSIDADE FEDERAL DA BAHIA  
INSTITUTO DE CIÊNCIA DA INFORMAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO**

**LEVI ALÃ NEVES DOS SANTOS**

**CONTRIBUIÇÃO DA MINERAÇÃO DE DADOS E DA OTIMIZAÇÃO  
HEURÍSTICA PARA A INTERPRETAÇÃO DOS DADOS DA  
PRODUÇÃO CIENTÍFICA BRASILEIRA**

Salvador  
2011

LEVI ALÃ NEVES DOS SANTOS

**CONTRIBUIÇÃO DA MINERAÇÃO DE DADOS E DA OTIMIZAÇÃO  
HEURÍSTICA PARA A INTERPRETAÇÃO DOS DADOS DA  
PRODUÇÃO CIENTÍFICA BRASILEIRA**

Dissertação apresentada ao Programa de Pós-Graduação do Instituto de Ciência da Informação da Universidade Federal da Bahia, como requisito parcial para obtenção do grau de Mestre.

Orientadora: Profa. Dra. Ana Paula de Oliveira Villalobos

Salvador  
2011

S237 Santos, Levi Alã Neves dos

Contribuição da mineração de dados e da otimização heurística para a interpretação dos dados da produção científica brasileira / Levi Alã Neves dos Santos. – Salvador, 2011.

114 f.; il. col.

Orientador: Profa. Dra. Ana Paula de Oliveira Villalobos  
Dissertação (mestrado) - Universidade Federal da Bahia. Instituto de Ciência da Informação, 2011.

1. Produção científica. 2. Mineração de dados. 3. Faces de Chernoff. 4. Análise de componentes principais. 5. Lógica Difusa. I. Universidade Federal da Bahia. II. Villalobos, Ana Paula de Oliveira. III. Título.

TERMO DE APROVAÇÃO

LEVI ALÃ NEVES DOS SANTOS

**CONTRIBUIÇÃO DA MINERAÇÃO DE DADOS E DA OTIMIZAÇÃO  
HEURÍSTICA PARA A INTERPRETAÇÃO DOS DADOS DA  
PRODUÇÃO CIENTÍFICA BRASILEIRA**

Dissertação aprovada como requisito parcial para obtenção do grau de mestre em Ciência da Informação, Universidade Federal da Bahia, pela seguinte banca examinadora:

**Ana Paula de Oliveira Villalobos** \_\_\_\_\_

Doutorado em Educação com interface em novas Tecnologias da Informação e da Comunicação, UFBA, Brasil

**Valter de Senna** \_\_\_\_\_

Doutorado em Pesquisa Operacional, University of Southampton, Inglaterra  
Pós-doutorado em Estatística, University of Southampton, Inglaterra

**Maria Teresa Navarro de Britto Matos** \_\_\_\_\_

Doutorado em Educação, UFBA, Brasil

Salvador, 25 de agosto de 2011

À Heitor e Arthur, meus amados filhos, os  
maiores presentes de minha vida.

## AGRADECIMENTO

À Deus que permite todas as coisas. Muito obrigado Senhor.

À meus pais, Martins e Tânia, e irmãos, Hales e Aline, por serem perseverantes em me mostrar o caminho seguro para a felicidade. Amo vocês.

À minha esposa, Joanna Amorim Suñé Neves, por segurar em minha mão sempre que acho que vou fraquejar me fazendo continuar e nunca desistir. Obrigado, te amo.

À minha eterna professora, Vanda Angélica da Cunha, por estar sempre presente nos momentos em que mais preciso. Mãezona, obrigado por toda orientação, apoio e puxões de orelha. Sou seu eterno fã.

À minha orientadora, Ana Paula de Oliveira Villalobos, por acreditar na importância e relevância desse estudo e incentivar seu desenvolvimento.

Ao professor Valter de Senna por plantar a primeira semente e acreditar na aplicabilidade desse estudo.

À professora Maria Teresa Navarro de Britto Matos, sempre compreensível e sensível às necessidades e possibilidades de seus orientandos. Muito obrigado por tudo.

À professora Henriette Ferreira Gomes, coordenadora do programa de pós-graduação. Brilhante mestre de cerimônia nessa maravilhosa festa da pesquisa em Ciência da Informação. Obrigado por sempre estar disposta a ouvir mesmo quando existe um mundo em suas costas.

À minha equipe da Politécnica, chefe atencioso, colegas compreensíveis e parceiros. Obrigado, vocês são 10<sup>3</sup>.

Aos professores, em especial à Teresinha Fróes, Nídia Lubisco, Graça Teixeira, Carmélia e Othon Jambeiro, por conduzirem um aprendizado, que vai além da técnica, impregnado de valores morais e éticos. Verdadeiros Mestres.

À toda equipe do ICI sempre muito dispostos a ajudar. Valeu por tudo.

*“Ao lado do cipreste branco  
À esquerda da entrada do inferno  
Está a fonte do esquecimento:  
Vou mais além, não bebo dessa água.  
Chego ao lago da memória  
Que tem água pura e fresca  
E digo aos guardiões da entrada:  
-Sou filho da Terra e do Céu.  
Dai-me de beber, que tenho uma sede sem fim.”*

*A Fonte, Legião Urbana*

## RESUMO

Estudo experimental com métodos de mineração de dados e de otimização heurística. Destaca a aplicação de mapas temáticos, faces de Chernoff e lógica difusa na análise de dados multivariados, possibilitando estabelecer classificação a partir de regras pré-estabelecidas, controladas. Os métodos asseguram base para planejamento, acompanhamento e análise da produção científica nacional. Tem por objetivo analisar os critérios de mensuração e apresentação dos dados sobre a produção científica nacional e descrever e aplicação da mineração de dados e da lógica difusa como métodos auxiliares na análise desses dados, através do contraste entre a demonstração de sua visualização e os métodos tradicionais. Desenvolve análise comparativa da produção por estado brasileiro e demonstra que tais métodos descrevem a produção com informações de impacto no fenômeno. Os resultados compõem técnicas de visualização de dados baseadas em representação visual de fácil associação e entendimento. Indicam a perspectiva de novas hipóteses de análise e aponta aspectos interessantes do perfil de produção brasileira. O modelo construído comprova a aplicabilidade do método, o que recomenda sua utilização pela Ciência da Informação.

**Palavras-chave:** produção científica; mineração de dados; faces de Chernoff; lógica difusa.



## **ABSTRACT**

Experimental study using methods of the data mining and heuristic optimization. Proof use of thematic maps, Chernoff faces and diffuse logic in multivariate data analysis, enabling classification set from predefined rules, controlled. The methods provide the framework for planning, monitoring and analysis of scientific production. The objective is to analyze the methods of measurement and display of data on scientific production and the description and application of data mining and diffuse logic as auxiliary methods in data analysis, through the contrast between the view presented and traditional methods. Develops a comparative analysis of the Brazilian production and states that describe the methods with information of impact. The results comprise data visualization techniques based on visual representation of association and easy understanding. Indicate the prospect of new hypotheses and points of interesting aspects analysis of the profile of Brazilian production. Proof the application of the method to model.

**Keywords:** scientific indicators, data mining, Chernoff faces, diffuse logic.

## LISTA DE TABELAS

TABELA 6.3.1- PRODUÇÃO BIBLIOGRÁFICA SEGUNDO UNIDADES DA FEDERAÇÃO DE PESQUISADORES DOUTORES, 2005-2008, CENSO 2008.....	41
TABELA 6.3.2 - ÍNDICES <i>PER CAPITA</i> DA PRODUÇÃO CIENTÍFICA A PARTIR DO NÚMERO DE DOUTORES .....	49
TABELA 9.1- CONTRIBUIÇÕES DAS VARIÁVEIS (%) EM RELAÇÃO AS COMPONENTES F1 E F2 .....	60

## LISTA DE QUADROS

QUADRO 6.3.1 – RELAÇÃO DOS INDICADORES COM AS VARIÁVEIS .....	50
QUADRO 10.1.1 – DEFINIÇÃO DAS VARIÁVEIS DA FACE .....	68
QUADRO 11.1.1 - PROJETO ESTATÍSTICO DIFUSO .....	77
QUADRO 11.1.2 – VARIÁVEIS DE ENTRADA.....	78
QUADRO 11.1.3 – VARIÁVEL DE SAÍDA.....	80
QUADRO 11.1.4 - BASE DE REGRAS DIFUSAS .....	81
QUADRO 11.1.5 - RESULTADO DIFUSO.....	82

## LISTA DE FIGURAS

FIGURA 6.2.1 - CORRELAÇÃO POSITIVA $R > 0$ .....	39
FIGURA 6.2.2 - CORRELAÇÃO NEGATIVA $R < 0$ .....	39
FIGURA 6.2.3 - CORRELAÇÃO NÃO-LINEAR.....	39
FIGURA 6.2.4 - SEM CORRELAÇÃO $R = 0$ .....	39
FIGURA 7.1.1 - ARTIGOS COMPLETOS PUBLICADOS EM PERIÓDICOS ESPECIALIZADOS DE CIRCULAÇÃO NACIONAL.....	53
FIGURA 7.1.2 - ARTIGOS COMPLETOS PUBLICADOS EM PERIÓDICOS ESPECIALIZADOS DE CIRCULAÇÃO INTERNACIONAL.....	54
FIGURA 7.1.3 - TRABALHOS COMPLETOS PUBLICADOS EM ANAIS DE EVENTOS.....	55
FIGURA 7.1.4 - LIVROS PUBLICADOS.....	56
FIGURA 7.1.5 - CAPÍTULOS DE LIVRO PUBLICADOS.....	56
FIGURA 7.1.6 - OUTRAS PUBLICAÇÕES BIBLIOGRÁFICAS.....	57
FIGURA 7.1.7 - RESUMOS DE TRABALHOS PUBLICADOS EM PERIÓDICOS ESPECIALIZADOS ..	58
FIGURA 7.1.8 - RESUMOS DE TRABALHOS PUBLICADOS EM ANAIS DE EVENTOS.....	58
FIGURA 10.1 - EXEMPLOS DE GLIFOS MULTIVARIADA SEGUNDO WARD (2008).....	66
FIGURA 10.1.1 - FACES DE CHERNOFF DA PRODUÇÃO CIENTÍFICA DOS ESTADOS BRASILEIROS.....	69
FIGURA 10.1.2 - FACES DE CHERNOFF DA PRODUÇÃO CIENTÍFICA DOS ESTADOS BRASILEIROS COM IDENTIFICAÇÃO DE GRUPOS.....	70
FIGURA 11.1.1 - ESTRUTURA DO SISTEMA DIFUSO.....	76
FIGURA 11.1.2 - VARIÁVEIS DIFUSAS PARA ARTIGOS INTERNACIONAIS - ARTINT.....	78
FIGURA 11.1.3 - VARIÁVEIS DIFUSAS PARA ARTIGOS NACIONAIS - ARTNAC.....	79
FIGURA 11.1.4 - VARIÁVEIS DIFUSAS PARA TRABALHO EM ANAIS - TRACOM.....	79
FIGURA 11.1.5 - VARIÁVEIS DIFUSAS PARA CLASSIFICAÇÃO DOS ESTADOS - FINAL.....	79
FIGURA 11.1.6 - MAPA DA PRODUÇÃO DE ARTIGOS.....	83

## LISTA DE GRÁFICOS

GRÁFICO 6.3.1 - DIAGRAMA DE DISPERSÃO ARTIGOS PUBLICADOS EM PERIÓDICOS NACIONAIS X DOUTORES.....	42
GRÁFICO 6.3.2 - DIAGRAMA DE DISPERSÃO ARTIGOS PUBLICADOS EM PERIÓDICOS NACIONAIS SEM O ESTADO DE SÃO PAULO X DOUTORES.....	43
GRÁFICO 6.3.3 - DIAGRAMA DE DISPERSÃO ARTIGOS PUBLICADOS EM PERIÓDICOS INTERNACIONAIS X DOUTORES .....	44
GRÁFICO 6.3.4- DIAGRAMA DE DISPERSÃO TRABALHOS COMPLETOS PUBLICADOS EM ANAIS DE EVENTOS X DOUTORES.....	44
GRÁFICO 6.3.5 - DIAGRAMA DE DISPERSÃO LIVROS PUBLICADOS X DOUTORES.....	45
GRÁFICO 6.3.6 - DIAGRAMA DE DISPERSÃO CAPÍTULOS DE LIVRO PUBLICADOS X DOUTORES .....	45
GRÁFICO 6.3.7 - DIAGRAMA DE DISPERSÃO OUTRAS PUBLICAÇÕES BIBLIOGRÁFICAS X DOUTORES .....	45
GRÁFICO 6.3.8 - DIAGRAMA DE DISPERSÃO RESUMOS DE TRABALHOS PUBLICADOS EM PERIÓDICOS ESPECIALIZADOS X DOUTORES .....	45
GRÁFICO 6.3.9 – DIAGRAMA DE DISPERSÃO RESUMOS DE TRABALHOS PUBLICADOS EM ANAIS DE EVENTOS X DOUTORES.....	46
GRÁFICO 6.3.10 - HISTOGRAMA DO TOTAL DE PRODUÇÃO CIENTÍFICA POR ESTADO.....	47
GRÁFICO 6.3.11– HISTOGRAMA DE ÍNDICE <i>PER CAPITA</i> DA PRODUÇÃO CIENTÍFICA POR NÚMERO DE DOUTORES .....	48
GRÁFICO 9.1 - SCREE PLOT DAS COMPONENTES PRINCIPAIS.....	61
GRÁFICO 9.2 – ACP DAS VARIÁVEIS.....	62
GRÁFICO 9.3 – ACP DAS OBSERVAÇÕES .....	63
GRÁFICO 11.1.1 - CLASSIFICAÇÃO DA PRODUÇÃO DE ARTIGOS .....	83

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>16</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b> .....	<b>20</b>
<b>3</b>	<b>A CIÊNCIA DA INFORMAÇÃO E A MENSURAÇÃO DO CONHECIMENTO</b> ..	<b>22</b>
<b>4</b>	<b>USO DE INDICADORES PARA TOMADA DE DECISÃO</b> .....	<b>28</b>
<b>5</b>	<b>APRESENTAÇÃO DOS DADOS</b> .....	<b>34</b>
<b>6</b>	<b>ANÁLISE DE DADOS UNIVARIADOS</b> .....	<b>36</b>
6.1	A APLICAÇÃO DE FERRAMENTAS ESTATÍSTICAS .....	36
6.2	CORRELAÇÃO E REGRESSÃO LINEAR .....	37
6.3	VERIFICANDO A CORRELAÇÃO ENTRE AS VARIÁVEIS.....	40
<b>7</b>	<b>VISUALIZANDO OS DADOS ATRAVÉS DE MAPAS TEMÁTICOS</b> .....	<b>51</b>
7.1	ÍNDICES <i>PER CAPITA</i> ATRAVÉS DE MAPAS TEMÁTICOS .....	52
<b>8</b>	<b>ANÁLISE DE DADOS MULTIVARIADOS</b> .....	<b>59</b>
<b>9</b>	<b>ANÁLISE DE COMPONENTES PRINCIPAIS</b> .....	<b>60</b>
<b>10</b>	<b>TÉCNICAS ICONOGRÁFICAS</b> .....	<b>65</b>
10.1A	APLICAÇÃO DAS FACES DE CHERNOFF AOS DADOS .....	68
<b>11</b>	<b>A LÓGICA DIFUSA</b> .....	<b>72</b>
11.1	A APLICAÇÃO DA LÓGICA FUZZY AOS DADOS DA PRODUÇÃO CIENTÍFICA .....	75
<b>12</b>	<b>CONSIDERAÇÕES FINAIS</b> .....	<b>85</b>
<b>13</b>	<b>REFERÊNCIAS</b> .....	<b>87</b>
<b>APÊNDICES</b> .....		<b>91</b>
APÊNDICE 1 – RELATÓRIO GERADO COM OS DADOS DA ANÁLISE DOS COMPONENTES PRINCIPAIS DO SOFTWARE XLSTAT .....		92
APÊNDICE 2 – RELATÓRIO GERADO COM OS DADOS DE ANÁLISE DA LÓGICA FUZZY DO SOFTWARE FUZZYTECH .....		100
1	GENERAL INFORMATION .....	100
TABLE OF CONTENTS .....		100

LIST OF FIGURES .....	100
LIST OF TABLES .....	100
LIST OF ABBREVIATIONS .....	100
FUZZY_FINAL.....	100
PROJECT DESCRIPTION .....	101
SYSTEM STRUCTURE .....	101
VARIABLES .....	101
INPUTS .....	102
OUTPUTS .....	103
INPUT VARIABLE "ARTINT" .....	103
INPUT VARIABLE "ARTNAC" .....	104
INPUT VARIABLE "TRACOM" .....	104
OUTPUT VARIABLE "FINAL" .....	105
RULE BLOCKS .....	106
RULE BLOCK "RB1" .....	106
SETTINGS.....	107
<b>ANEXOS .....</b>	<b>108</b>
ANEXO 1 – PORTAL DO PLANO TABULAR.....	109
ANEXO 2 – VERSÃO ANTERIOR DE ACESSO AOS INDICADORES POR UNIDADE DA FEDERAÇÃO NO PORTAL DO MCT .....	112

## 1 INTRODUÇÃO

No universo das bibliotecas e de outros centros de informação, percebe-se que o grande volume de informação é sempre tema de discussão. A responsabilidade do profissional bibliotecário é criar soluções para efetuar o controle e a disseminação do recurso de conhecimento através de sistemas eficientes e arquitetura de informação eficaz. As Tecnologias de Informação e Comunicação (TICs) além de transmitir, também têm o papel de ajudar nos processos de reflexão, indicando para onde caminha o conhecimento.

Os indicadores de produção científica nacional são obtidos após uma árdua garimpagem em diversas bases e são apresentados à sociedade através de tabelas e gráficos em valores quantitativos de produtos e produtores, como é descrito pelo próprio Ministério da Ciência e Tecnologia, MCT. Esse método empregado pelo MCT apresenta pelo menos dois pontos a serem melhorados: os estados são representados de forma igualitária e a não utilização de métodos estatísticos que considerem a possível inter-relação entre os indicadores da produção científica.

Há de se convir que determinar produtividade sem definir parâmetros e modelos de análise é muito difícil, embora para muitos seja possível considerar e retratar a realidade brasileira através de colunas, linhas e barras comuns nos estudos de dados com base em planilhas eletrônicas. Aparentemente o que o MCT faz é dispor de dados brutos e carentes de uma análise mais detalhada que permita uma avaliação informativo-descritiva. Ferramentas da mineração de dados e da otimização heurística podem ampliar o escopo de análise de dados feita por ciências como a bibliometria e a ciência da informação e trazer mais critérios e confiabilidade aos resultados.

Para discutir a análise da produção científica nacional este estudo recorre aos fundamentos da análise de dados multivariados. Tal método é muito utilizado no desenvolvimento de novas tecnologias no campo das ciências exatas e biológicas, o que com toda certeza têm muito a contribuir para o aperfeiçoamento e a valorização das análises obtidas a partir de estudos bibliométricos e cienciométricos, sempre preocupados com o estudo do desenvolvimento científico e da produção científica.



A proposta de pesquisa aponta para a valorização da possibilidade de recorrer aos conhecimentos da matemática e da estatística para ampliar a área da compreensão de dados sobre a produção científica nacional.

A dissertação está estruturada em 12 capítulos delineados de acordo com os elementos considerados como necessários a formulação e execução da atividade de estudo acadêmico. Os capítulos encontram-se articulados do ponto de vista conceitual e da argumentação proposta pelo autor.

Na introdução o destaque está no campo conceitual, apresentando a questão pesquisa, levanta a hipótese da sua atividade de pesquisa, define os objetivos gerais e específicos, refere-se à metodologia do procedimento de pesquisa.

Os segundo, terceiro e quarto capítulos se constituem da apresentação dos fundamentos teóricos identificados durante a revisão de literatura que deverão subsidiar a análise dos dados coletados a cerca do objeto estudado.

O quinto e parte do sexto capítulos descreve a construção do quadro metodológico. Descrevendo os procedimentos e apresentando instrumentos da atividade de pesquisa. Ainda no sexto capítulo expõem a coleta de dados e procede a análise dos dados coletados de acordo com a interpretação do pesquisador apoiado na revisão de literatura.

As análises dos dados que se estende do sétimo ao décimo primeiro capítulos, com uma evolução natural e espontânea do tema, procurando ampliar a perspectiva da pesquisa.

O décimo segundo capítulo corresponde às considerações finais onde o autor tenta retomar o que desenvolveu nos capítulos anteriores, com recomendação para utilização do modelo pela Ciência da Informação como uma nova proposta de abordagem para análise dos dados.

Embora hoje, a mineração de dados, ainda seja o método mais utilizado nas ciências exatas, quando empregadas pela Ciência da Informação em seus estudos proporcionará um notável alcance na possibilidade de transformar informação em conhecimento. Entendendo que a análise de dados estatísticos permite uma maior segurança na tomada de decisões.

Dentro do tema de estudos métodos para interpretação dos indicadores de produção científica brasileira surgiu a **questão de pesquisa** que considerava como a mineração de dados e a lógica difusa podem promover uma melhor compreensão dos dados sobre a produção científica nacional. O desenvolvimento da pesquisa procurou confirmar a **hipótese** de que existe uma relação entre a produção científica e o número de doutores envolvidos nas pesquisas e que a produção científica nacional é, numericamente, equivalente entre os estados.

O estudo do tema e a busca de resposta à questão de pesquisa justificam-se pelo fato de que nos tempos atuais, Ciência e Tecnologia são elementos cruciais de transformação e desenvolvimento para as nações, por isso, definir indicadores significativos e entendê-los melhor tornam os investimentos nessa área mais eficientes e eficazes.

O **objetivo geral** visou analisar os critérios de mensuração e apresentação dos dados sobre a produção científica nacional e descrever a utilização da mineração de dados e a lógica difusa como métodos auxiliares na análise desses dados, através do contraste entre sua demonstração de visualização de dados e os métodos tradicionais. Para alcançá-lo traçou-se como **objetivos específicos** apontar os atuais critérios de análise da produção científica brasileira utilizada na tomada de decisão e avaliação dos estados produtores; descrever a utilização dos métodos mineração de dados e lógica difusa na análise da produção científica; contrastar a visualização dos dados a partir da aplicação dos métodos de mineração de dados e lógica difusa com os métodos tradicionais.

Nesse sentido, a construção de instrumentos, que dêem maior suporte à avaliação e às decisões das agências de fomento, torna-se imprescindível para racionalizar e flexibilizar tanto a aplicação de recursos públicos quanto a definição de políticas nos estados e na Federação.

Do ponto de vista da natureza é uma pesquisa aplicada descritiva com abordagem quantitativa, mas com possível compreensão qualitativa a partir das análises cognitivas existentes nas apresentações visuais dos dados.

Os procedimentos são baseados no estudo de caso, tendo como base o método dedutivo. Para validar a pesquisa foi requerida a existência de correlação

entre a produção científica dos estados e o número de doutores envolvidos nas pesquisas, utilizando os dados da produção do ano de 2008 apresentados pelo MCT.

A pesquisa envolve a utilização de um índice relativo ou *per capita* para a visualização da real produtividade dos estados com base na quantidade de doutores. Utilizando o índice relativo foram aplicados os métodos de mineração na visualização dos dados: mapas temáticos e Faces de Chernoff.

Por fim, a aplicação da lógica difusa para classificar o perfil produtivo dos estados brasileiros e a apresentação de seus resultados através de gráfico setorial e mapa temático.

## 2 FUNDAMENTAÇÃO TEÓRICA

A fundamentação teórica utilizada para análise baseia-se, principalmente, na visão interdisciplinar da Ciência da Informação apresentada por Borko (1968). Ciência da Informação é a disciplina que investiga as propriedades e o comportamento da informação, as forças que regem o fluxo de informações e os meios de processamento para melhor acessibilidade e usabilidade. Também derivada e interrelacionada com outras áreas, dentre elas a Matemática e Biblioteconomia fundamentais para este estudo.

Pelo menos dois pontos dos cinco indicados por Borko (1968), são relevantes para a Ciência da Informação, pois justificam esta pesquisa. O primeiro é sobre o grande número de cientistas trabalhando e o grande número de revistas científicas e técnicas que existem hoje e o segundo é o curto período entre a investigação e a aplicação que torna a necessidade de informações mais premente e mais imediata.

As apreensões de Wersig (1993) afirmam que além de interdisciplinar o trabalho teórico na Ciência da Informação, tem que ser interconceitual para atingir algum alcance, podendo ser considerado como: evolucionista, sinóptico e transdisciplinar. O autor aponta que o principal problema seria o campo de estudo da Ciência da Informação como objeto de muitas disciplinas fragmentadas, tendo que lidar com todos estes itens fracionados de uma natureza empírica ou teórica. O novo tipo de ciência não está previamente dirigido pela busca de um entendimento completo de como nosso mundo funciona mas pela necessidade de lidar com problemas e resolvê-los.

Autores explicam, como Saracevic (1995) que a Ciência da Informação no aspecto da recuperação da informação, contribui para esse estudo por apontar que hoje, para as práticas profissionais de negócios com base em gestão da informação e do conhecimento, há mais interdisciplinaridade em todos os esforços de Pesquisa e Desenvolvimento (P&D).

Como resultado da explosão da informação, exemplificada com a enorme propagação da Internet e os conceitos de infra-estrutura global de informação, o número de campos e os novos atores sociais estão se movendo para lidar com a informação.

Para a Ciência da Informação essas pressões estão provocando, entre outros, cooperações mais interdisciplinares. Pelo perfil de aplicabilidade dos métodos expostos na inteligência competitiva e inter-relacionar com a análise da produção científica este estudo utiliza o entendimento sobre o que são dados, informação e conhecimento sugeridos por Davenport (1998).

Por fim, foram aplicados os estudos didáticos de Moita Neto (2004) sobre análise univariável e multivariável de dados que permitem aos pesquisadores, de campos de pesquisas não familiarizados aos cálculos estatísticos, beneficiarem-se do grande impacto que esses métodos trazem ao desenvolvimento de estudos científicos. Reforça o papel das tecnologias nesse momento de grande desenvolvimento científico, apontando uso de ferramentas que possibilitem um novo olhar sobre dados dizendo que talvez neste momento, tenha chegado a um paradoxo interessante, à complexidade matemática pode ser substituída por uma simplicidade didática.

Para a pesquisa as trivialidades estatísticas — incluindo a análise multivariada — estão ao alcance de todos e sem o constrangimento matemático do passado. Através do uso de software estatístico, é possível pensar estatística sem ser estatístico (MOITA NETO, 2004).

### 3 A CIÊNCIA DA INFORMAÇÃO E A MENSURAÇÃO DO CONHECIMENTO

Durante toda sua história, o homem foi fascinado por compreender todas as coisas que o rodeavam. Desejo registrado em todas as culturas que existem ou já existiram no planeta. Toda essa necessidade, ou melhor, desejo de saber inicialmente ficou conhecida como filosofia e com o passar do tempo essa busca incessante ficou conhecida como ciência.

Não que a filosofia abandonasse seu desejo pelo saber mas o foco desse desejo precisou ser demonstrado e ter métodos de possível experimentação caso existissem dúvidas sobre as demonstrações requeridas. Discussões a parte sobre o escopo de abrangência desses modelos de abordagem do pensamento humano, o que interessa nesse estudo é apresentar uma linha de raciocínio que surge com a necessidade de preservar o conhecimento, manter este registro atualizado e tentar prever o futuro.

Nos primeiros momentos de registro do conhecimento existia apenas a preocupação em mantê-lo a salvo dos inimigos, esses poderiam ser qualquer um. É fácil encontrar relatos de autores como Wilson Martins, que em sua obra *A Palavra Escrita*, aborda o tema e até mesmo em obras de ficção amplamente conhecidas como o *Nome da Rosa* de Umberto Eco.

A informação era privilégio dos eruditos e estava retida pelos muros dos mosteiros cuidada e vigiada pelos monges (BARRETO 2008). Nesse período o conhecimento estava encerrado nas páginas de um livro, seja este em pergaminho, papiro e até mesmo em tábuas de argila. Todo conhecimento estava represado em páginas de um determinado suporte físico. O livro era responsável, focando aqui na unicidade do termo, pela disseminação do conhecimento e sua possível censura. Cada livro era um mundo... controlável.

O surgimento da imprensa determinou um novo papel para o livro, o suporte era o mesmo, mas a forma de divulgar o pensamento ganhou um forte aliado: a duplicação. Todo o ambiente da época colaborou para que aquele desejo reprimido explodisse com toda força. É um período áureo para o saber. Barreto (2008) argumenta que esse momento realmente só aconteceu nos dias atuais explicando que

[...] a informação esteve cativa em universos simbólicos divinos por longos anos. Entre alforrias e prisões chegou até a época da Internet onde grande parte dos textos é liberado completo em sua linguagem natural. Mas muitos insistem em continuar operando por uma sublime recapitulação do passado.

Antes da invenção das tecnologias de impressão, o conhecimento era de certa forma um saber pessoal e organizado na forma oral tradicional. Embora com a invenção dos sistemas de conhecimento escrito pudesse ser armazenado a qualquer momento, sua disseminação dependia em parte que as pessoas fossem capazes de ler e transferir o conhecimento oralmente para outra pessoa.

A maior invenção de Gutenberg não foi a imprensa, mas uma tecnologia que permitiu que mais pessoas pudessem escrever seus conhecimentos pessoais e apresentá-los a outras pessoas (WERSIG, 1993). É justamente nesse período que os alicerces das ciências mais antigas têm suas fundações aprofundadas e surgem também as novas ciências. A ciência dos livros tem seu reconhecimento e a Biblioteconomia aparece para o mundo assumindo sua preocupação com o livro e deste resolve cuidar.

Muitos estudos nessa área têm suas leis escritas e estruturadas. A descrição e a indexação são pensadas para cercar, no sentido de limitar, esse controle ao livro. As principais estruturas lógicas da organização de bibliotecas, levando em consideração seu acervo, seu público e seu desenvolvimento também datam desse período. Por mais que isso seja questionado, seu real papel social começa a ser desenhado nesse período. O renascimento realmente foi para tudo e todos. Foi também nesse período que nomes considerados visionários da área de Biblioteconomia apontam para uma ciência nova que trate dos meandros do objeto nebuloso: a informação.

Percebeu-se que o conhecimento até então depositados nos livros deveriam ser visto como potencial fonte de novos conhecimentos. Observaram algo que parece óbvio, mas que ainda hoje é difícil de ser compreendido por alguns, os livros se inter-relacionam, eles conversam entre si ou são complementos uns dos outros. O suporte não era mais relevante e sim o que suas páginas continham e como seria possível inter-relacionar esses conteúdos.

Em 1895 Paul Otlet e La Fontaine se destacam com a concepção do livro enquanto documento, apontando em uma nova direção, mas ainda com a preocupação criar meios de acesso ao suporte com o objetivo de preservar o conhecimento.

Pesquisadores apontam que a partir desses estudos destinados agora ao documento se inicia uma ciência nova que, mais tarde será conhecida como Ciência da Informação. Muitos afirmam que o foco não está mais no livro nem em seu suporte, mas agora apenas no conteúdo (conhecimento), que está sob o olhar de estudos. Esse cenário começa a mudar com a introdução de um novo conceito de informação que até os dias atuais pode ser considerado um assunto complexo e multidimensional, embrionário de uma nova ciência, que tem tanto um componente de ciência pura, que investiga o assunto sem considerar a sua aplicação e um componente de ciência aplicada, que desenvolve serviços e produtos (BORKO, 1968).

Às vezes tem-se a impressão que existem mais publicações paradigmáticas no mercado do que relatórios de pesquisas substanciais de considerável seriedade. Se olharmos criticamente para estas publicações parece que as discussões paradigmáticas não mostram ser causadas por mudança paradigmática dramática (porque o trabalho científico real segue em frente como antes), nem por sérias competições entre paradigmas alternativos (WERSIG, 1993).

Nasce a Ciência da Informação com a concepção de contribuir para que as pesquisas tenham a liberdade de passear por onde for conveniente na busca de produzir conhecimento. Embora houvesse um longo caminho a percorrer, a CI teve seu início no começo do século XX com o advento da documentação como sendo um modo (não muito) prático de lidar com algo que era sentido como um problema.

A Ciência da Informação é interdisciplinar na natureza. Devido a isto as relações com várias disciplinas estão mudando e sua evolução interdisciplinar está longe de terminar (SARACEVIC, 1995).

O grande questionamento ainda está na abordagem na qual existem outras tradições dentro da ciência da informação, que não se ajustam na estrutura da biblioteca ou da recuperação da informação, tais como por exemplo: estudos do



fluxo da informação, estudos das conseqüências sociais das tecnologias da informação e estudos sobre a produção do conhecimento (WERSIG, 1993).

Segundo Borko (1968), a Ciência da Informação é uma ciência interdisciplinar derivada e relacionada com áreas como matemática, lógica, linguística, psicologia, informática, pesquisa de operações, artes gráficas, comunicação, biblioteconomia, administração e outras áreas afins (BORKO, 1968).

Pode se observar que, através de muitos séculos o papel do conhecimento para indivíduos, organizações e sociedades mudou de várias formas e estas mudanças se tornaram aparentes no começo do século passado, aproximadamente desde a década de 1960, estão se tornando parte de uma ampla mudança que às vezes chamamos de pós-modernismo (WERSIG, 1993).

A filosofia desses estudos interdisciplinares continuam de alguma forma a proposta de Otlet e La Fontaine no século XIX e posteriormente a de Vannevar Brush com seu artigo Como nós pensamos (As we may think), em 1945, de criar um planetário universal que aponte qualquer conhecimento produzido pelo homem em qualquer tempo e lugar.

Sendo a Ciência da Informação uma participante ativa na evolução dessa chamada sociedade da informação (SARACEVIC 1995), autores como Borko (1968) e Wersig (1993) colocam luz sobre as discussões quando apontam para um novo olhar dado às ciências novas.

Naquele tempo com o fenômeno da explosão da informação, surgiram questões complexas e não poderiam ser respondidas compreensivelmente. Os autores explicam que os possíveis questionamentos vêm de outro modelo, já em novo cenário, a visão ocidental de determinar fronteiras e de estudos multidisciplinares perde força frente ao crescente entendimento interdisciplinar das ciências. As ciências não só trabalham juntas, mas elas estão inter-relacionadas e contribuem para o mesmo fim de produzir conhecimento.

A Ciência da Informação é um campo dedicado à investigação científica e à prática profissional, com a finalidade de resolver os problemas da comunicação efetiva dos registros do conhecimento em atendimento às necessidades de

informação, sendo inexoravelmente ligada à tecnologia da informação (SARACEVIC, 1995). Este fenômeno não está restrito ao conhecimento científico (WERSIG, 1993).

O termo informação nunca aparece sozinho onde quer que ele esteja. Isso pode ser observado em qualquer área de estudo que necessite do termo para definir algo que contribui para o ato de conhecer. Isso explica porque o termo parece ser definido sempre de forma genérica. Não existe um consenso entre os vários autores sobre o conceito de informação.

Tomando-se a necessidade de se fazer distinção entre *dados*, *informação* e *conhecimento*, Davenport (1998) explica como é difícil definir o termo informação por entender que quaisquer distinções são nitidamente imprecisas e que além do mais, informação é um termo que envolve todos os três. Para este, dados são observações sobre o estado do mundo; informação é o mesmo que, dados dotados de relevância e propósito; e conhecimento, é na verdade a informação mais valiosa da mente humana.

Também tendemos a exagerar o significado dessas palavras. Durante anos, as pessoas se referiram a *dados* como *informação*; agora, vêem-se obrigadas a lançar mão de *conhecimento* para falar sobre a informação — daí a popularidade da 'administração do conhecimento'. (DAVENPORT, 1998, p. 18)

A informação definida no contexto da produção do conhecimento gera mais informação é para muitos um ciclo interminável e para alguns um ciclo controlável. Não importa em que área esse conhecimento é produzido e nem onde as informações oriundas desse conhecimento vão chegar. Não existem fronteiras para a atuação da Ciência da Informação e nem limitações em seus métodos porque o conhecimento é universal, ou seja, está em todas as outras ciências.

Borko (1968) afirma que na Ciência da Informação há espaço tanto para o campo teórico quanto para o prático e, claramente, ambos são necessários. Teoria e prática estão inexoravelmente ligadas, cada uma alimenta-se do trabalho da outra.

Sem as limitações que a Biblioteconomia encontrou por ter seu objeto de investigação no suporte da informação, a Ciência da Informação vem proporcionando ambientes de desenvolvimento úteis às outras ciências. Para muitos esta é a diferença fundamental entre Biblioteconomia e Ciência da

Informação. A primeira se consolidou criando todo um cenário que privilegiava as bibliotecas com a visão de apropriação do conhecimento, muito mais pelo reconhecimento de sua importância histórica do que pelo seu potencial econômico. A segunda apropriando-se da informação como seu objeto.

A era do conhecimento, como é tratado esse período recente da história, tem sua significação promovida pelo reconhecimento dos aspectos econômicos que a informação alcançou. Diversas correntes de pensamentos convergem para a necessidade do domínio dos mecanismos que fazem parte do controle da informação. A Ciência da Informação emprega estudos, métodos e conceitos das demais áreas do conhecimento, o que a mantém sempre jovem e dinâmica. Isso não é um problema epistemológico e sim um novo conceito de ciência (WERSIG, 1993).

O dinamismo apresentado pela Ciência da Informação é fundamental para que ela atue em campos distintos e tente solucionar os problemas informacionais com abrangência. Essa é uma ciência que possibilita a atuação de vários profissionais das mais diversas áreas, que tentam resolver seus problemas informacionais utilizando a versatilidade e já alguma experiência da Ciência da Informação, CI. Esses atores sociais agregam as informações oriundas de determinado conhecimento de sua área a fim de criar novos conhecimentos, tendo como mediadora a CI. Por isso os métodos e fundamentos da CI muitas vezes têm como base outras áreas do conhecimento, mas o objeto é sempre o mesmo: a informação.

Como indicou Borko (1968), as técnicas e procedimentos utilizados pelos bibliotecários e documentalistas são, ou deveriam ser, concebidos a partir do campo teórico da Ciência da Informação. A Ciência da Informação como uma disciplina tem objetivo de fornecer um conjunto de informações, as quais conduzam ao desenvolvimento das técnicas associadas ao ciclo informacional.

#### 4 USO DE INDICADORES PARA TOMADA DE DECISÃO

Com o avanço das novas tecnologias, é possível observar alterações nos chamados tradicionais setores da economia. A informação, enquanto produto, ao lado da agropecuária, indústria e serviços vêm se formando o que se pode chamar de quarto setor, passando assim a sociedade a dar tal valor à informação bem como a outros recursos naturais, favorecendo o desenvolvimento social (BURNHAM, 2005).

Ao longo dos anos o ser humano vem aprimorando o processo de preservação do conhecimento e do desenvolvimento, através do acesso e uso da informação. O nível de (in)formação da população pode ser avaliado pelo seu desenvolvimento tecnológico e de acesso à informação (BURNHAM, 2005). O Índice de Desenvolvimento Econômico (IDH) aponta o nível de acesso à informação, viabilizado pelas Tecnologias da Informação e da Comunicação, como uma variável determinante para o desenvolvimento social (RUAS, 1998), (BOLAÑO; MELO, 2000).

Atualmente, as organizações já assimilaram a idéia que a informação precisa e em tempo hábil é crucial para o desenvolvimento de planos estratégicos e para a própria gestão do empreendimento.

Choo (2003) divide o uso estratégico do conhecimento em 3 arenas<sup>1</sup> e aponta para a dificuldade de aprendizado dentro das organizações e suas conseqüências. Para Choo (2003, p. 28), os atores sociais nas organizações acham difícil reexaminar crenças herdadas e questionar as práticas existentes como as únicas alternativas viáveis.

Uma grande massa de informação deve ser processada antes de ser transformada em conhecimento. Portanto, cada vez mais estamos necessitando de ferramentas estatísticas que apresentem uma visão mais global do fenômeno que aquela possível numa abordagem univariada. (MOITA NETO, 2004)

Dados brutos tomados como informação processada certamente prejudicam a caminhada em direção ao conhecimento. Embora a tomada de decisão seja um

---

<sup>1</sup> Arenas: termo empregado por Choo (2003) para identificar as formas de utilização da informação estratégica dentro das organizações.

processo complexo, não há dúvida de que ela é uma parte essencial da vida da organização. A ação na empresa é provocada por uma decisão e toda decisão é um compromisso para uma ação (CHOO, 2003, p. 29).

A informação é um componente intrínseco de quase tudo que uma organização faz. Sem uma clara compreensão dos processos organizacionais e humanos pelos quais a informação se transforma em percepção, conhecimento e ação, as empresas não são capazes de perceber a importância de suas fontes e tecnologias de informação (CHOO, 2003, p. 27).

A informação deve proporcionar um novo olhar dentro de um determinado contexto, sendo que seu potencial é complexo e não simplesmente linear. Castells (1999) afirma que o ponto central desse debate é o uso da informação para gerar conhecimento, não de maneira linear, mas de forma cíclica onde a realimentação criaria uma espiral de inovação e usos.

No contexto de competitividade, baseada na contínua evolução de tecnologias e na crescente modernização das práticas administrativas, verifica-se a necessidade da utilização de métodos capazes de captar o que existe em termos de métodos para entender, adaptar e modificar de modo a ajustar os referidos avanços às condições concretas de cada cenário observado.

Uma das preocupações crescente no mundo, no que tange ao desenvolvimento científico e tecnológico é a de medir a produção científica. No Brasil, o Ministério da Ciência e Tecnologia (MCT), vem adotando, há algumas décadas, várias medidas para implementar tais mecanismos que gerenciem o controle do processo produtivo retratando o cenário nacional. Essa preocupação não só com o desenvolvimento nacional, mas também em tornar o país mais competitivo no mercado mundial é sem dúvida baseada na compreensão dos dados dessa produção (FAPESP, 2005).

Através do uso de novas tecnologias é possível efetivar as políticas públicas para o acesso e disseminação da informação. O ideal não é só participar das decisões no que tange a esse universo, mas também contribuir para a construção de projetos políticos relacionados à democratização da informação.

Os números apontam um aumento significativo de publicações nacionais nos últimos anos, com destaque para o crescimento dessa participação em bases internacionais importantes como a *Web of Science*.

Durante muito tempo, o MCT estuda as melhores formas de traduzir esse crescimento e também as melhores formas de investir em ciência e tecnologia. Desenvolver políticas voltadas para o setor é um dos principais objetivos dos estudos, subsidiando com dados os órgãos de fomento que, por sua vez, utilizam-se desses na escolha da melhor forma de incentivar o desenvolvimento científico e tecnológico do país.

Sobretudo, não se pode esquecer a responsabilidade das universidades, que mesmo dentro de uma visão social e de busca pela democratização da informação, têm que estar dispostas e convictas de que o uso de novas tecnologias interferiram no seu papel e que o conhecimento nunca foi tão fundamental dentro de uma sociedade quanto em uma sociedade globalizada. Instituições como universidades e centros acadêmicos, por sua vez, devem oferecer condições para ampliar a criação e a expansão do conhecimento. A relação com a sociedade implica na construção de uma cultura democrática rompendo com o autoritarismo social que se apresenta enraizado em nossa cultura.

A preocupação em entender as relações entre as atividades científicas e suas relações com o desenvolvimento econômico e social tem elevado a importância da análise dos indicadores de produção científica. Esses indicadores são apresentados como principal ferramenta para compreender a dinâmica dos processos científicos e fornecer elementos para direcionar políticas em ciência e tecnologia, os indicadores de produção científica são cada vez mais necessários ao planejamento e avaliação de resultados.

Segundo a FAPESP (2005, p. 5) os indicadores de produção científica, somados à família de indicadores de insumos para a Ciência e Tecnologia (C&T), têm contribuído de forma definitiva para a análise do desempenho e melhoria da eficiência dos sistemas nacionais de ciência, tecnologia e inovação.

Os indicadores podem ser compreendidos como dados estatísticos usados para medir algo intangível, que ilustram aspectos de uma realidade multifacetada. A construção e o uso de indicadores de

produção científica são objeto de estudo de várias áreas do conhecimento, sendo usados tanto para o planejamento e a execução de políticas para o setor como também para que a comunidade científica conheça melhor o sistema no qual está inserida. (FAPESP, 2005, p 5.)

No Brasil, foi a partir dos anos 80 que surgiram os primeiros esforços para gerar indicadores de C&T através do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ), do próprio Ministério da Ciência e Tecnologia (MCT), com a coleta e a publicação de informações sobre os recursos do governo federal aplicados em C&T, utilizando as recomendações da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) para os gastos em P&D, e as sugestões da UNESCO, para as atividades científicas e técnicas correlatas – ACTC (BRASIL, 2010).

No campo específico da produção científica foi o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) quem deu os primeiros passos na produção de indicadores já na década de 70. Já em 1999, O Ministério da Ciência e Tecnologia (MCT), assumiu a responsabilidade pela organização e a divulgação das informações de C&T do País, de forma centralizada através da ajuda de instituições públicas, no âmbito federal e estadual, além de organizações privadas (BRASIL, 2010).

Os indicadores elaborados e mantidos pela Ministério da Ciência e Tecnologia (MCT) seguem as recomendações dos Manuais da chamada "Família Frascati" da Organização para a Cooperação e Desenvolvimento Econômico (OCDE). Organização intergovernamental que tem a preocupação eminentemente voltada para o desempenho econômico e para a ação política objetiva a alcançar uma melhor compreensão das relações entre a C&T e a economia (PORCARO, 2005).

A Organização para a Cooperação e Desenvolvimento Econômico (OCDE) é responsável pelo conjunto de documentos que inclui manuais específicos para a área de P&D (Manual Frascati), inovação (Manual de Oslo), balanço de pagamentos tecnológico (Manual TBP), recursos humanos (Manual de Canberra) e ainda o Manual de Patentes utilizados pelo MCT.

Para o Ministério da Ciência e Tecnologia (MCT) a adoção de padrões internacionais não é conflitante com o objetivo das instituições nacionais. O

importante é que o sistema de informação seja capaz de gerar respostas ágeis e adequadas às necessidades de planejamento, acompanhamento e avaliação das respectivas bases científicas e tecnológicas. Afirma ainda que iniciativas regionais, como as desenvolvidas no âmbito da Rede Ibero-Americana de Indicadores de Ciência e Tecnologia (RICYT), podem levar também ao estabelecimento de um conjunto padronizado de indicadores, melhor adaptado às particularidades da região, sem implicar, necessariamente, no rompimento com as recomendações internacionais (BRASIL, 2010).

Além de estabelecer recomendações e regras para contabilizar as atividades de P&D, a OCDE publica uma série padronizada de indicadores de insumos e resultados para o seu grupo de países, tendo se tornado referência básica para outras iniciativas nacionais. Embora esses indicadores não estejam isentos de críticas, constituem, sem nenhuma dúvida, a base comum para os vários países na geração de séries históricas e comparáveis de informação (BRASIL, 2010).

O MCT assume o papel de fornecer informações especializadas à comunidade científica, aos formuladores de políticas, aos legisladores, à imprensa, aos alunos e à sociedade em geral. Afirma que essa ainda que mera tentativa de apreensão de uma realidade complexa, os indicadores de ciência e tecnologia permitem vislumbrar um pouco do país que teremos no futuro.

Vale ressaltar o esforço do governo e da sociedade em apresentar o retrato do estágio no domínio do conhecimento científico e tecnológico e a direção do desenvolvimento social e econômico do país.

Mesmo diante dessa responsabilidade o MCT informa que dos indicadores de resultados apenas os desenvolvidos a partir da Bibliometria apresentaram resultados mais avançados e criteriosos. O Ministério explica que são ainda incipientes as tentativas de elaboração de indicadores de impacto, isto é, medir como determinado resultado científico ou tecnológico afeta as várias dimensões das condições de existência dos indivíduos, seja no próprio campo científico e tecnológico, na dimensão econômica e na dimensão social.

Para o MCT toda a problemática da construção e seleção de indicadores se mantém através do constante aperfeiçoamento e depuração. Esse processo inclui também o uso de novas tecnologias de medição e aferição com base em dois



pontos importantes: a representação das transformações no domínio da ciência, da tecnologia e da inovação e o acompanhamento dos padrões metodológicos internacionais de elaboração de indicadores.

## 5 APRESENTAÇÃO DOS DADOS

Este estudo apresenta métodos de análise de dados da produção científica brasileira utilizando informação da tabela de produção bibliográfica segundo unidades da Federação para pesquisadores doutores, no período 2005-2008, referentes ao Censo 2008. Essas informações são originadas do Plano Tabular do CNPq e são a base para o relatório do Ministério de Ciência e Tecnologia.

Todas as observações partem da premissa de que a apresentação dessas informações de produção, nos veículos de divulgação e a sua posterior análise não retratam a realidade e singularidade de estados e regiões do país, uma vez que, o formato de divulgação utilizado na apresentação destes dados através de tabelas e gráficos se faz útil, mas simples demais para uma análise mais aprofundada da real produção científica nacional.

O Portal do Plano Tabular<sup>2</sup> tem a finalidade de divulgar o perfil da pesquisa no Brasil em termos quantitativos, organizando esses dados em tabelas configuradas dinamicamente pelo usuário. Esse sistema possibilita o cruzamento de variáveis capazes de gerar um número muito grande de diferentes tabelas, que podem ser salvas em planilhas ou em arquivos de texto para futuras consultas.

Conforme informação do Portal, o Plano Tabular se apóia em sete conjuntos básicos de dados (unidades de análise) compostos pelas informações relativas aos Grupos de Pesquisa, aos Pesquisadores, Estudantes, ao Pessoal Técnico, às Linhas de Pesquisa, à interação com o Setor Produtivo e à Produção Científica, Tecnológica e Artística. Da conjunção dessas informações, originadas do cadastramento dos grupos de pesquisa, da base de currículos *Lattes* e do Sistema Coleta/Capes, foram corrigidas quaisquer inconsistências próprias de uma captura de dados. Sobre a base corrigida construiu-se um conjunto de tabelas de agregados e fatos que facilitam e aumentam o desempenho das consultas.

Ainda segundo o Portal, o inventário da produção científica, tecnológica e artística dos grupos foi construído a partir das informações existentes nos currículos

---

<sup>2</sup> Disponível em: <<http://dgp.cnpq.br/planotabular/>>

Lattes dos pesquisadores e estudantes. Como consequência, quase sempre haverá duplas ou múltiplas contagens no número de produções. Em outras palavras, a produção CT&A dos grupos é sempre apresentada por uma *proxi*, que é a soma das produções individuais de seus componentes. Pesquisadores que participam de mais de um grupo de pesquisa terão a totalidade de sua produção remetida a cada um dos grupos de que participa.

Esse estudo teve como base a análise dos seguintes indicadores de produção científica disponibilizado pelo MCT: artigos completos publicados em periódicos especializados de circulação nacional, artigos completos publicados em periódicos especializados de circulação internacional, trabalhos completos publicados em anais de eventos, livros publicados, capítulos de livro publicados, outras publicações bibliográficas, resumos de trabalhos publicados em periódicos especializados, resumos de trabalhos publicados em anais de eventos. Em 2006, esses resultados foram apresentados diretamente na página do MCT que podem ser observados nos anexos desse trabalho.

Baseado em técnicas de visualização de dados a partir da aplicação de técnicas de análise multivariada de dados, o estudo tem duas abordagens: técnicas univariadas para comparar a produção dos Estados e técnicas de análise multivariadas para obter novas comparações. Entre os resultados obtidos destaca-se a semelhança dos esforços produtivos nacionais.

## 6 ANÁLISE DE DADOS UNIVARIADOS

Moita Neto (2004) explica que os dados quantitativos permitem uma avaliação estatística mais rica e podem ser sempre transformados em dados qualitativos. Portanto, uma pergunta sobre renda, por exemplo, pode ser apresentada como dado numérico (valor da renda), dado ordinal (faixa de renda) ou dado nominal (rico ou pobre).

Ainda segundo Moita Neto (2004), estabelecer relações, encontrar ou propor leis explicativas é o papel próprio da ciência. Para isso é necessário controlar, manipular, medir as variáveis que são consideradas relevantes ao entendimento do fenômeno analisado. Muitas são as dificuldades em traduzir as informações obtidas em conhecimento. Porém, no caso restrito de variáveis independentes entre si é possível, com razoável segurança, interpretar um fenômeno complexo usando as informações estatísticas de poucas variáveis. As informações estatísticas mais relevantes neste tipo de análise são as medidas de tendência central e de dispersão dos dados.

### 6.1 A APLICAÇÃO DE FERRAMENTAS ESTATÍSTICAS

As explicações que se seguem foram obtidas a partir de observações e orientações didáticas dos trabalhos de Martins e Donaire (1981), Farias (2002) e Martins (2005) e de informações e fórmulas disponíveis na Wikipédia. Os trabalhos analisados tinham por objetivo orientar não estatísticos a utilizarem ferramentas estatísticas para desenvolver pesquisas e observações dadas em pesquisas em qualquer área do conhecimento.

Martins (2005) explica que diariamente é apresentada uma grande quantidade de informações numéricas e dependendo das situações assim surge a necessidade de conhecimentos e capacitações para compreender essas informações. Os procedimentos, técnicas e métodos estatísticos são fundamentais para auxiliarem na execução da tarefa de compreender informações numéricas produzidas por outros (*data users*), bem como construir informações numéricas a partir de dados brutos (*data producers*). Esse autor também afirma que as facilidades proporcionadas pelo uso de *softwares* estatísticos têm estimulado o uso de tais procedimentos nos mais diversos campos de pesquisa.

## 6.2 CORRELAÇÃO E REGRESSÃO LINEAR

O termo correlação significa relação em dois sentidos (co+relação), e é usado em estatística para designar a força que interrelacionam dois conjuntos de valores. A verificação da existência e do grau de relação entre as variáveis é o objeto de estudo da correlação. Quando é encontrado certo grau de relacionamento, verificado pela correlação entre duas variáveis, aplica-se a análise de regressão que vai permitir descrever através de um modelo matemático, essa relação, partindo de  $n$  observações das mesmas. Para executarmos a regressão, as variáveis serão divididas em variável dependente e variável independente. Para o eixo  $x$ , indicamos a variável independente e para o eixo  $y$ , a dependente. Em termos gerais, a correlação mede a força, ou grau, de relacionamento entre duas variáveis; a regressão dá uma equação que descreve o relacionamento em termos matemáticos.

O coeficiente de correlação  $\rho_{X,Y}$  entre duas variáveis aleatórias  $X$  e  $Y$  com valores esperados  $\mu_X$  e  $\mu_Y$  e desvios padrão  $\sigma_X$  e  $\sigma_Y$  é definida como:

$$\rho_{X,Y} = \text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y},$$

onde  $E$  é o operador valor esperado e  $\text{cov}$  significa covariância. Como  $\mu_X = E(X)$ ,  $\sigma_X^2 = E(X^2) - E^2(X)$  e, do mesmo modo para  $Y$ , podemos escrever também

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y},$$

O primeiro passo de análise deste estudo foi comprovar a hipótese inicial da existência de relação entre a variável número de doutores com as demais variáveis indicativas de produção científica, isto é, saber se as variações quantitativas presentes em uma das variáveis são acompanhadas por alterações nas outras. Por exemplo, número de doutores em relação à produção de artigos em revistas nacionais, ou em relação à produção de artigos em revistas internacionais, ou em relação à produção de livros, etc.

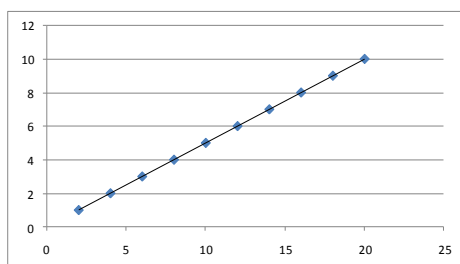
Uma medida do grau e do sinal da correlação é dada pela covariância entre as duas variáveis aleatórias  $X$  e  $Y$ , que é uma medida numérica de associação linear

existente entre as variáveis, ou seja, covariância ou variância conjunta é uma medida de como duas variáveis variam conjuntamente.

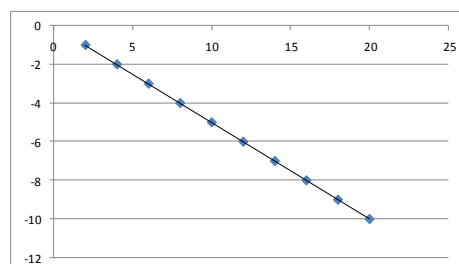
Nas observações da relação entre o variável número de doutores com as demais variáveis indicativas de produção científica foi utilizado o coeficiente de correlação linear de Pearson ( $\rho$  de Pearson), além de ser o mais conhecido, mede o grau da correlação e mostra a direção (positiva ou negativa) dessa correlação. Este coeficiente, normalmente representado por  $\rho$  assume apenas valores entre -1 e 1. Quando  $\rho = 1$  Significa uma correlação perfeita positiva entre as duas variáveis. Quando  $\rho = -1$  Significa uma correlação negativa perfeita entre as duas variáveis - Isto é, se uma aumenta, a outra sempre diminui. E caso  $\rho = 0$  Significa que as duas variáveis não dependem linearmente uma da outra. No entanto, pode existir uma dependência não linear. Assim, o resultado  $\rho = 0$  deve ser investigado por outros meios. O coeficiente de correlação linear de Pearson é obtido através da seguinte equação:

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \cdot \text{var}(Y)}}$$

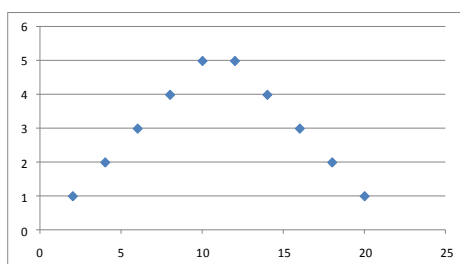
O diagrama de dispersão mostrará que a correlação será tanto mais forte quanto mais próximo estiver o coeficiente de  $-1$  ou  $+1$ , e será tanto mais fraca quanto mais próximo o coeficiente estiver de zero. Abaixo aparecem figuras representativas criadas nesse estudo para ilustrar os modelos de correlação.



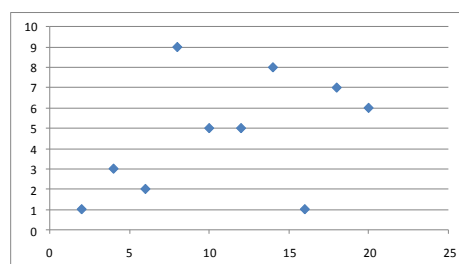
**Figura 6.2.1** - Correlação positiva  $r > 0$



**Figura 6.2.2** - Correlação negativa  $r < 0$



**Figura 6.2.3** - Correlação não-linear



**Figura 6.2.4** - sem correlação  $r = 0$

Os gráficos apresentados nas observações da correlação entre o número de doutores e as de produção científica neste estudo informam também o coeficiente de determinação  $R^2$  que indica a proporção de variação da variável independente que é explicada pela variável dependente, ou seja, é uma ferramenta que avalia a qualidade do ajuste. Quanto mais próximo da unidade o  $R^2$  estiver, melhor a qualidade do ajuste. O seu valor fornece a proporção da variável Y explicada pela variável X através da função ajustada. Exemplo:  $R^2 = (0,9929)^2 = 0,9858 = 98,50 \%$ .

Para finalizar esse entendimento preliminar é necessário lembrar algumas observações. Primeiro, correlação não é o mesmo que causa e efeito, isso implica que duas variáveis podem estar altamente correlacionadas e, no entanto, não haver relação de causa e efeito entre elas. Segundo, o estudo de correlação pressupõe que as variáveis X e Y tenham uma distribuição normal. E por fim, o coeficiente de determinação  $R^2$  é o responsável por indicar a qualidade da correlação.

### 6.3 VERIFICANDO A CORRELAÇÃO ENTRE AS VARIÁVEIS

Para entender a relação entre as variáveis número de doutores e dados de produção científica apresentados da Tabela 6.3.1, o procedimento para a análise da correlação foi aplicado aos pares gerando o gráfico da reta de regressão e o coeficiente de correlação.

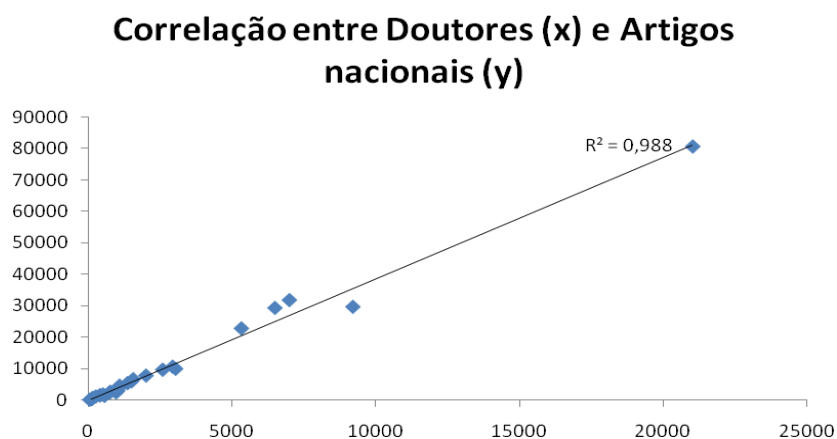
Na análise das duas variáveis o objetivo é conhecer se entre estas existe alguma possível dependência, isto é, se valores altos/baixos de uma das variáveis estão correlacionados com os valores altos/baixos da outra. Uma correlação alta (próxima de 1) pode indicar forte dependência linear entre as variáveis. Nesse caso, os pontos no diagrama de dispersão espalham-se em torno de uma reta. No diagrama, a variável, números de doutores, sempre será representada pelo eixo x (horizontal) e as demais se revezaram no eixo y (vertical).



**Tabela 6.3.1- Produção bibliográfica segundo unidades da Federação de pesquisadores doutores, 2005-2008, Censo 2008**

UF	Total de Doutores	Artigos completos publicados em periódicos especializados		Trabalhos completos publicados em anais de eventos	Livros ou capítulos de livro publicados		Outras publicações bibliográficas (3)	Resumos de trabalhos publicados em (4)	
		Nacional	Internacional		Livros	Capítulos de livros		Periódicos especializados	Anais de eventos
Acre	125	320	174	234	23	179	698	3	808
Alagoas	567	1344	1149	2817	160	835	1820	87	4137
Amapá	54	98	154	94	22	33	256	6	458
Amazonas	966	2418	2848	2836	326	1578	4410	136	8124
Bahia	3040	9935	7672	11215	896	5225	14234	501	22142
Ceará	1504	5960	5324	7711	434	2910	7621	290	14814
Distrito Federal	2009	7761	6202	9507	942	4885	12926	305	16098
Espírito Santo	753	2422	1632	4248	209	1545	3584	65	5810
Goiás	1367	5359	3721	6111	437	2619	8457	266	10865
Maranhão	410	1502	1491	1316	115	527	1582	157	3992
Mato Grosso	766	2623	1265	2646	276	1056	3456	73	6154
Mato Grosso do Sul	1088	4559	2425	5280	385	2124	8206	114	8978
Minas Gerais	6992	31778	23117	35499	2433	12552	37260	1770	66079
Pará	1023	2925	2470	4353	329	1995	4682	74	7737
Paraíba	1565	6624	3789	12449	430	2650	7078	260	11308
Paraná	5324	22782	15431	29746	1728	8341	29212	1224	47301
Pernambuco	2588	9636	6779	13963	730	4868	13633	520	19524
Piauí	406	1501	1008	1293	111	675	2902	108	3257
Rio de Janeiro	9200	29640	34190	42092	3400	18503	37513	1565	64942
Rio Grande do Norte	1035	3596	2548	6251	302	1544	4469	205	9450
Rio Grande do Sul	6491	29292	22705	37713	2502	15622	35713	1443	63013
Rondônia	116	319	209	400	35	233	650	4	595
Roraima	154	686	313	329	39	222	1160	5	1287
Santa Catarina	2933	10615	7642	19607	1159	4978	12657	308	20764
São Paulo	21014	80603	91929	96572	6830	45787	107170	8090	218998
Sergipe	515	1786	1332	3107	219	792	2318	26	5577
Tocantins	266	1135	608	800	91	278	1785	35	1709

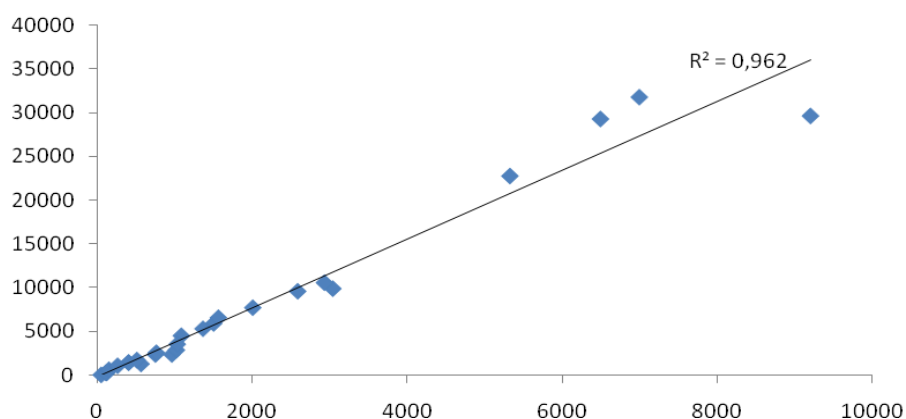
Na primeira análise, que utilizou os dados sobre artigos científicos publicados em revistas nacionais, foi obtida uma reta de regressão com os pontos distribuídos praticamente ao longo da reta apresentando forte associação entre as variáveis. O diagrama de dispersão gerado a partir dos dados apresenta a existência de relacionamento entre as variáveis, com altos valores de uma das variáveis associados a altos valores da outra variável, correlação positiva. O coeficiente de correlação, que marcou 0,992, também indica existir uma forte correlação entre a variável e o número de doutores observado no Gráfico 6.3.1. O  $R^2$  explica que 98,80% da variação do número de artigos nacionais é explicado pelo número de doutores em cada Estado.



**Gráfico 6.3.1** - Diagrama de dispersão artigos publicados em periódicos nacionais X doutores

O ponto que aparece no alto do gráfico em destaque representa o estado de São Paulo, em números absolutos é o maior produtor e o maior em concentração de doutores do país. Para dissipar qualquer dúvida sobre um possível erro ou desvio causado pela distância do ponto gerado pelos dados do estado de São Paulo, a análise refeita omitindo esses dados para verificar se de alguma forma São Paulo estava influenciando na análise da correlação, uma vez que este aparece isolado.

### Correlação entre Doutores (x) e Artigos nacionais (y) sem SP

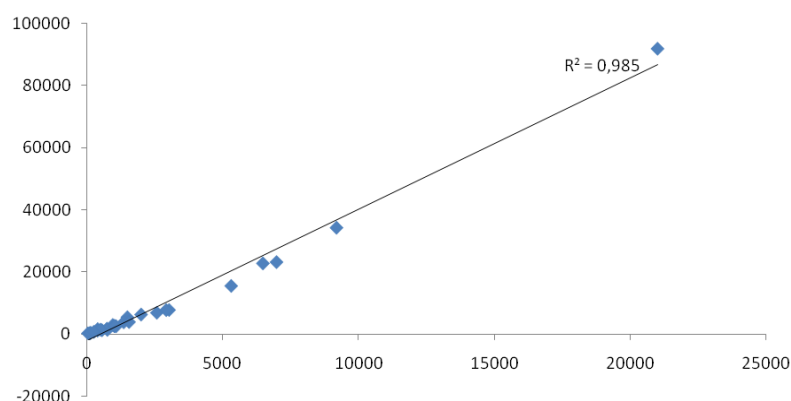


**Gráfico 6.3.2** - Diagrama de dispersão artigos publicados em periódicos nacionais sem o estado de São Paulo X doutores

Com a segunda análise foi possível constatar que realmente existe correlação entre o número de doutores e a produção de artigos nacionais com uma explicação de variação de 96,20%. Essa nova análise demonstrou que os dados do estado de São Paulo não estavam influenciando o resultado. Para uma maior segurança das futuras análises todas as demais variáveis de produção sofreram a mesma observação sem os dados do estado de São Paulo e o resultado obtido foi positivo para uma correlação entre o número de doutores e a produção científica.

A partir desse ponto, foi aplicada a análise de correlação para todas as demais variáveis da Tabela 6.3.2 com a segurança de que seus dados não sofriram efeitos discrepantes ao utilizar dados do estado de São Paulo. Os resultados obtidos apresentaram altos índices de correlação entre as variáveis. O destaque ficou por conta da publicação de livros com coeficiente de correlação de 0,996. Seguem os gráficos obtidos nas demais análises.

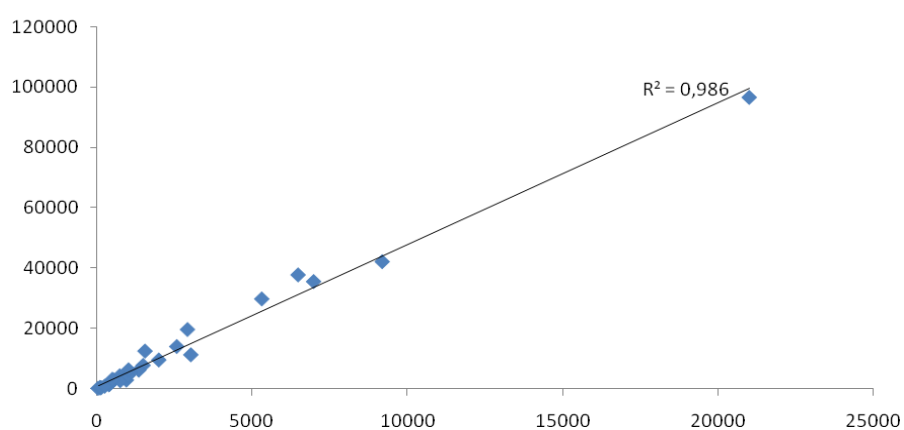
### Correlação entre Doutores (x) e Artigos internacionais (y)



**Gráfico 6.3.3** - Diagrama de dispersão artigos publicados em periódicos internacionais X doutores

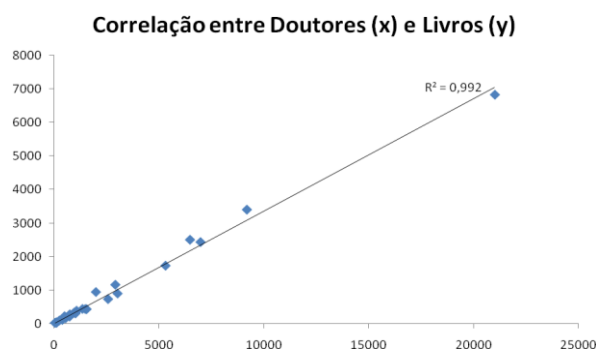
Os dados da correlação entre o número de doutores e a produção de artigos em periódicos internacionais apresentou um  $R^2$  de 0,985, isso explica que 98,5% da variação dos dados da produção são explicados pelo número de doutores envolvidos nesse processo. A forte correlação positiva demonstra que a qualidade da relação entre os dados possibilita dizer que existe dependência linear entre essas.

### Correlação entre Doutores (x) e Trabalhos em Anais (y)

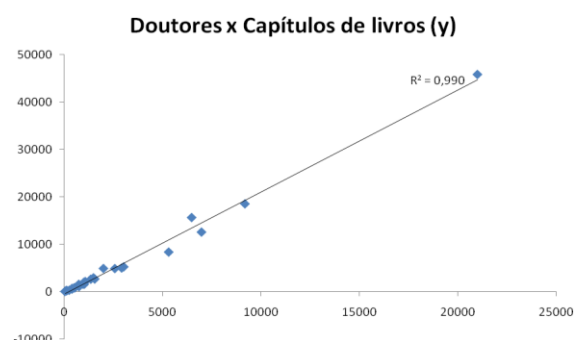


**Gráfico 6.3.4**- Diagrama de dispersão trabalhos completos publicados em anais de eventos X doutores

Os dados da produção de trabalhos completos em anais de eventos com correlação de 0,993 e  $R^2$  de 98,6% também indicam que os dados da produção científica tem forte correlação ao números de doutores em cada estado produtor. Os gráficos seguintes ajudam a identificar nos demais indicadores de produção científica a forte correlação positiva já apontada nas primeiras análises de correlação.

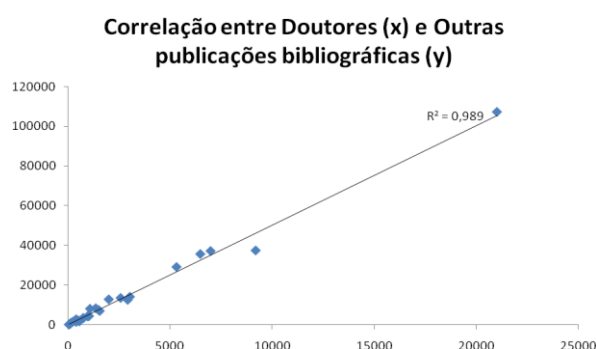


**Gráfico 6.3.5** - Diagrama de dispersão livros publicados X doutores

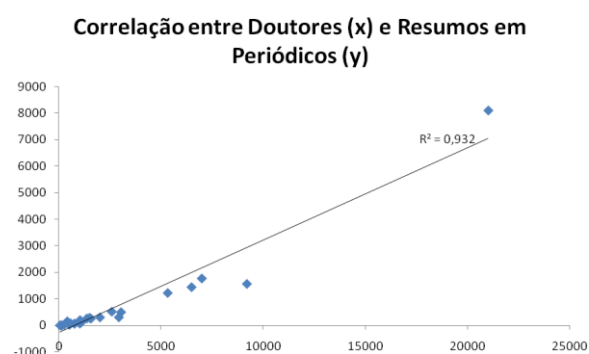


**Gráfico 6.3.6** - Diagrama de dispersão capítulos de livro publicados X doutores

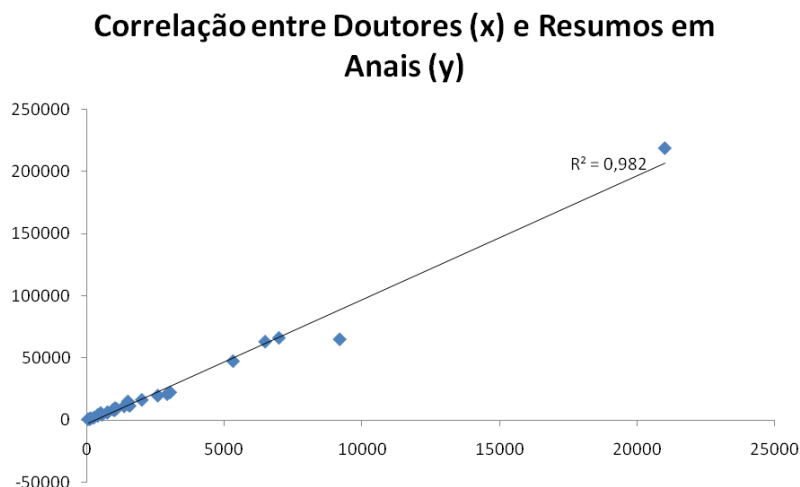
Produção de livros correlação de 0,996 e  $R^2$  de 99,2% e Capítulos de livros com correlação de 0,995 e  $R^2$  de 99% de relação na variação com o número de doutores.



**Gráfico 6.3.7** - Diagrama de dispersão outras publicações bibliográficas X doutores



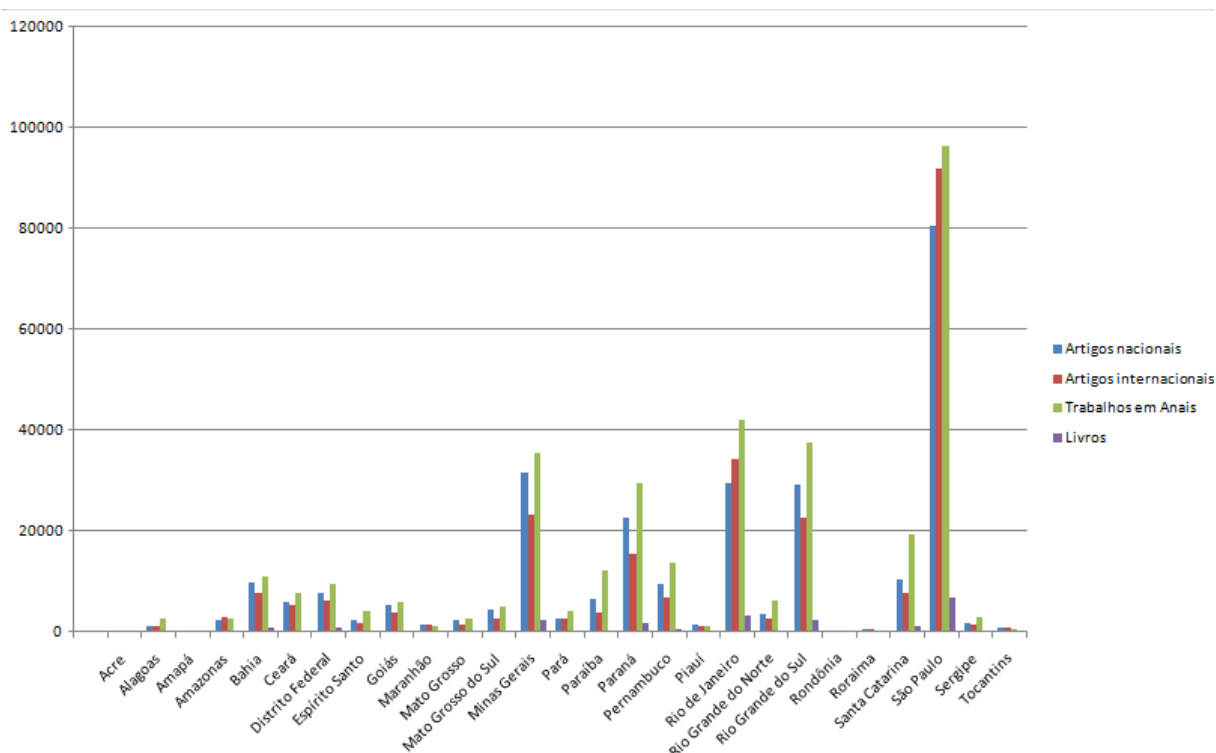
**Gráfico 6.3.8** - Diagrama de dispersão resumos de trabalhos publicados em periódicos especializados X doutores



**Gráfico 6.3.9** – Diagrama de dispersão resumos de trabalhos publicados em anais de eventos X doutores

Lembrando, correlação não indica uma causalidade e sim se existe relação entre duas variáveis. O uso dessa técnica se justificou primeiro para conhecer a relação entre produção e produtores. E segundo para servir de base à construção de um índice que relacione esse dois indicadores. Esse novo índice foi importante para determinar um dado quantitativo de produção proporcional ao número de doutores em cada estado, facilitando o entendimento dessa produção a partir de seus atores responsáveis por pesquisas fomentadas em todo o país.

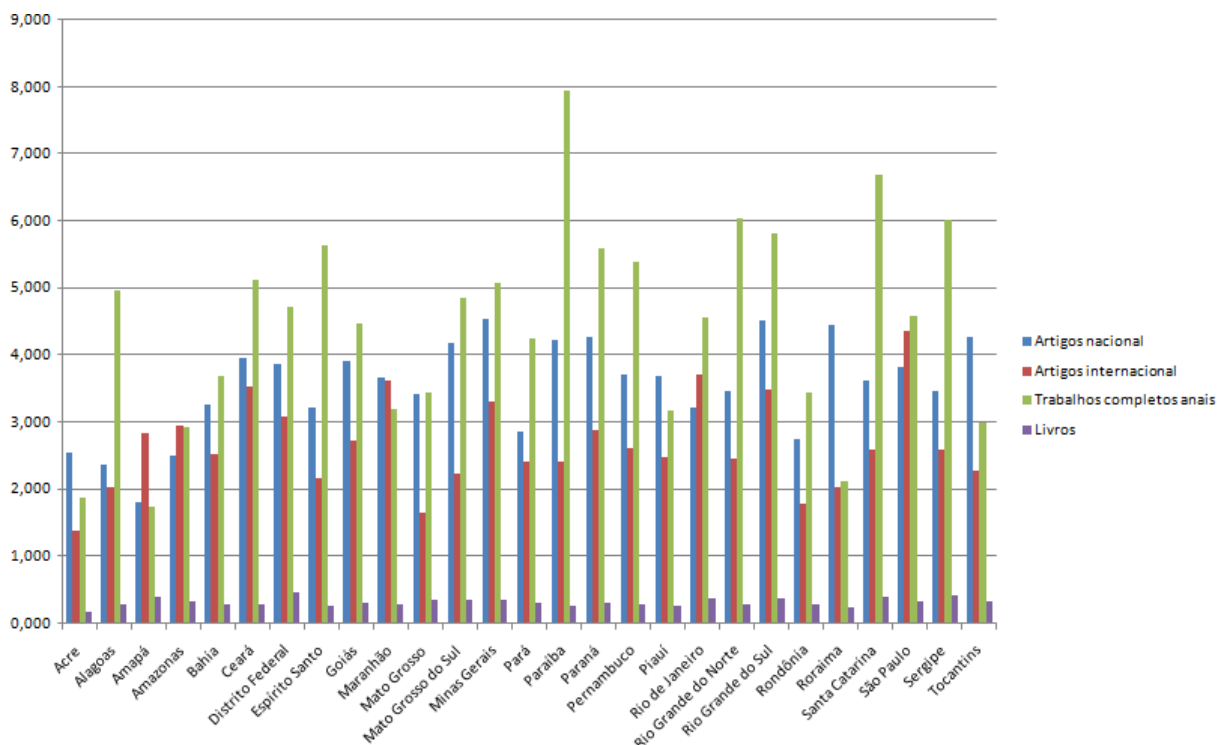
Para entender a diferença entre a utilização de outra técnica para interpretar os dados da produção dos estados, basta observar o Gráfico 6.3.10 e o Gráfico 6.3.11 que apresentam os dados de quatro dos indicadores. O Gráfico 6.3.10 exhibe os dados da produção científica a partir de seus indicadores como são apresentados pelo MCT e utilizados pelos órgãos de fomento em suas considerações.



**Gráfico 6.3.10** - Histograma do total de produção científica por estado

Já o Gráfico 6.3.11, Exibe esses mesmos dados mas considerando o número de doutores em cada estado. Essa dados fazem parte de um índice *per capita* onde para obtê-lo foram divididos os dados de produção científica pelo número de doutores em cada estado. *Per capita* é uma expressão latina que significa para cada cabeça para indicar uma média por pessoa de um dado valor (WIKIPÉDIA<sup>3</sup>).

<sup>3</sup> Wikipédia é uma enciclopédia multilíngue *online* livre colaborativa, ou seja, escrita internacionalmente por várias pessoas comuns de diversas regiões do mundo, todas elas voluntárias. Por ser livre, entende-se que qualquer artigo dessa obra pode ser transcrito, modificado e ampliado, desde que preservados os direitos de cópia e modificações, visto que o conteúdo da Wikipédia está sob a licença GNU/FDL (ou GFDL) e Creative Commons Attribution-ShareAlike (CC-by-SA) 3.0. Criada em 15 de janeiro de 2001, a Wikipédia baseia-se no sistema *wiki* (do havaiano *wiki-wiki* = "rápido", "veloz", "célere").



**Gráfico 6.3.11**– Histograma de índice *per capita* da produção científica por número de doutores

Examinando o primeiro desses dois gráficos observa-se a hegemonia absoluta do estado de São Paulo e que para alguns estados essa produção é tão ínfima que nem foi possível ser registrada. Já no segundo, o exame demonstra que as produções dos estados são quase que equivalentes. A Paraíba aparece hegemônica na produção de trabalhos em anais de eventos e São Paulo só supera os demais estados na produção de artigos em periódicos internacionais, mas acompanhado de perto pelo Ceará, Maranhão, Rio de Janeiro e Rio Grande do Sul. Foi possível também conhecer a produção do Acre, Amapá e Rondônia.

Para dar continuidade à análise dos dados foi desenvolvida a Tabela 6.3.2. Esta tabela contém os índices (*per capita*) relativos à produção (valores dos indicadores de produção científica dividida pelo número de doutores) de cada estado. A partir desta tabela foram realizadas todas as análises e estudos da aplicabilidade dos métodos de mineração de dados e otimização heurística escolhidos. Cada etapa de desenvolvimento utilizou parte ou o total desses indicadores segundo critérios explicados oportunamente. Os resultados obtidos foram de livre experimentação, mas com observações interessantes sobre o perfil produtor dos estados brasileiros.



**Tabela 6.3.2 - Índices *per capita* da produção científica a partir do número de doutores**

UF	ARTNAC	ARTINT	TRACOM	Livros	CAPLIV	OUTPUB	RESPER	RESANA
Acre	2,560	1,392	1,872	0,184	1,432	5,584	0,024	6,464
Alagoas	2,370	2,026	4,968	0,282	1,473	3,210	0,153	7,296
Amapá	1,815	2,852	1,741	0,407	0,611	4,741	0,111	8,481
Amazonas	2,503	2,948	2,936	0,337	1,634	4,565	0,141	8,410
Bahia	3,268	2,524	3,689	0,295	1,719	4,682	0,165	7,284
Ceará	3,963	3,540	5,127	0,289	1,935	5,067	0,193	9,850
Distrito Federal	3,863	3,087	4,732	0,469	2,432	6,434	0,152	8,013
Espírito Santo	3,216	2,167	5,641	0,278	2,052	4,760	0,086	7,716
Goiás	3,920	2,722	4,470	0,320	1,916	6,187	0,195	7,948
Maranhão	3,663	3,637	3,210	0,280	1,285	3,859	0,383	9,737
Mato Grosso	3,424	1,651	3,454	0,360	1,379	4,512	0,095	8,034
Mato Grosso do Sul	4,190	2,229	4,853	0,354	1,952	7,542	0,105	8,252
Minas Gerais	4,545	3,306	5,077	0,348	1,795	5,329	0,253	9,451
Pará	2,859	2,414	4,255	0,322	1,950	4,577	0,072	7,563
Paraíba	4,233	2,421	7,955	0,275	1,693	4,523	0,166	7,226
Paraná	4,279	2,898	5,587	0,325	1,567	5,487	0,230	8,884
Pernambuco	3,723	2,619	5,395	0,282	1,881	5,268	0,201	7,544
Piauí	3,697	2,483	3,185	0,273	1,663	7,148	0,266	8,022
Rio de Janeiro	3,222	3,716	4,575	0,370	2,011	4,078	0,170	7,059
Rio Grande do Norte	3,474	2,462	6,040	0,292	1,492	4,318	0,198	9,130
Rio Grande do Sul	4,513	3,498	5,810	0,385	2,407	5,502	0,222	9,708
Rondônia	2,750	1,802	3,448	0,302	2,009	5,603	0,034	5,129
Roraima	4,455	2,032	2,136	0,253	1,442	7,532	0,032	8,357
Santa Catarina	3,619	2,606	6,685	0,395	1,697	4,315	0,105	7,079
São Paulo	3,836	4,375	4,596	0,325	2,179	5,100	0,385	10,422
Sergipe	3,468	2,586	6,033	0,425	1,538	4,501	0,050	10,829
Tocantins	4,267	2,286	3,008	0,342	1,045	6,711	0,132	6,425

A fim de facilitar as análises e o uso dos indicadores de produção propostos pelo MCT, este estudo passa a utilizar um nome de variável específico para cada indicador segundo o Quadro 6.3.1. Esses nomes de variáveis foram usados em todas as análises desta pesquisa não sendo necessária a diferenciação entre o indicador e a variável, uma vez que esta última é apenas uma representação da primeira.

<b>INDICADOR</b>	<b>Variação do nome do INDICADOR</b>	<b>VARIÁVEL</b>
Artigos completos publicados em periódicos especializados de circulação nacional	Artigos em periódicos nacional(is)	ARTNAC
Artigos completos publicados em periódicos especializados de circulação internacional	Artigos em periódicos internacional(is)	ARTINT
Trabalhos completos publicados em anais de eventos	Trabalhos completos em anais	TRACOM
Livros publicados	Livros publicados	LIVRO
Capítulos de livro publicados	Capítulos de livro publicados	CAPLIV
Outras publicações bibliográficas	Outras publicações	OUTPUB
Resumos de trabalhos publicados em periódicos especializados	Resumos em periódicos	RESPER
Resumos de trabalhos publicados em anais de eventos	Resumos em anais	RESANA

**Quadro 6.3.1** – Relação dos indicadores com as variáveis

## 7 VISUALIZANDO OS DADOS ATRAVÉS DE MAPAS TEMÁTICOS

A visualização dos dados a partir do Gráfico 6.3.11 foi muito útil para uma interpretação da produção científica dos estados através da ótica *per capita*, mas para complementar a análise dos dados será empregada a técnica de visualização de dados a partir de mapas temáticos.

Mapas temáticos são representações simplificadas de um ambiente condicionado a um determinado tema. Construir mapas propõe apresentar determinadas características e especificidades inerentes a um território. Produzir mapas temáticos que combinam informações traz grandes vantagens em relação aos sistemas tradicionais. De forma visual os dados brutos e/ou desconexos são transformados em imagens e informações de fácil entendimento e associação (CARVALHO, 2009).

A elaboração de mapas temáticos tem como objetivo principal a representação das informações a fim de uma melhor visualização e comunicação. Um facilitador no uso desse método é o de que os fenômenos representados podem ser oriundos de qualquer área do conhecimento, seja das disciplinas exatas ou humanas. Neste estudo, os mapas foram construídos utilizando os índices das variáveis de cada estado dentro de faixas de valores observando distribuições em forma de grupos.

Para desenvolver os mapas foi utilizado o *Gismaps Viewer* um *software* de geoprocessamento o qual permite gerar mapas temáticos e visualizar imagens de satélite. Nesse estudo, os mapas foram construídos utilizando os índices *per capita* da produção de cada estado a partir das distâncias entre os valores máximos e mínimos dos índices dividido por três. Essa divisão corresponde às três faixas de valores que definem as cores do tema dos mapas.

Pela proximidade das medidas encontradas nos índices as três faixas apresentaram informações na tentativa de obter alguma diferenciação entre os estados. Para entender melhor os mapas, basta observar a legenda com as cores e os valores das faixas em cada mapa. Foram também mantidas as três cores da legenda em todos os mapas para uma melhor visualização da relação entre os mapas nos diferentes indicadores.

## 7.1 ÍNDICES *PER CAPITA* ATRAVÉS DE MAPAS TEMÁTICOS

Os índices *per capita* expressam o produto da produção científica com base no número de doutores em cada estado. O índice referente a cada indicador corresponde ao valor proporcional da produção por estado. Ao criar esse índice houve uma ampliação na forma de avaliar e de comparar como cada estado produz e como este aparece no cenário nacional. Para entender melhor, por exemplo, basta dividir o valor da produção de artigos em periódicos nacionais do estado de São Paulo de 80603 por 21014, referente ao número de doutores, e encontra-se o índice *per capita* de 3,836 artigos em periódicos nacionais por doutor.




Para entender a produção observando os mapas foi necessário definir faixas levando em consideração os valores máximos e mínimos. Lembrando que o número de faixas (três) foi definido pelo autor como uma forma didática apenas para apresentar os dados. Essas faixas foram definidas para variar entre valores eqüidistantes para cada indicador de produção científica. Por exemplo, para o valor máximo encontrado em um indicador igual a 10,0 e mínimo igual a 1,0, será obtido para as três faixas a variação igual a 3 a partir do seguinte processo:

$(\text{valor máximo} - \text{valor mínimo})/3 \rightarrow$  determina distância entre faixas

Para o exemplo temos:

$(10-1)/3 \rightarrow$  distâncias entre as faixas igual a 3

Então o quadro das legendas apresentado será:

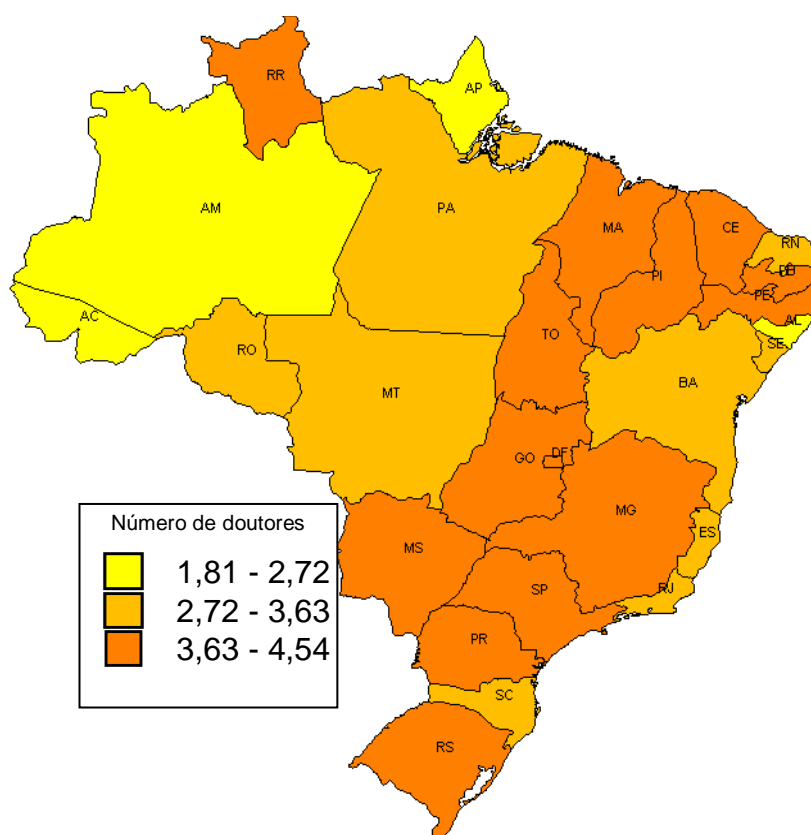
Número de doutores	
	1,0 – 4,0
	4,0 – 7,0
	7,0 – 10,0

Os índices encontrados demonstram uma grande similaridade na capacidade de produção de cada estado. Os valores orbitam uma faixa que em alguns casos atinge a distância entre o menor e o maior valor em menos de meio ponto, como é o caso da produção de livros e de resumos em periódicos. As maiores faixas são as apresentados por publicações de trabalhos completos em anais de eventos e

resumos em anais de eventos com uma variação de 6,2 e 5,7 pontos respectivamente. Existe um destaque também para o indicador outras publicações, mas por se tratar de um indicador para qualquer outra forma de publicação seus 4,3 pontos de faixa não demonstram ser expressivo.

O primeiro mapa temático, Figura 7.1.1, apresenta uma similaridade entre a produção científica dos estados. Os índices estão muito próximos em uma faixa de 0,91 pontos entre estas produções, sendo possível visualizar a formação de dois grupos maiores. Todas as regiões apresentam algum estado na faixa mais produtiva. A região Sul, por exemplo, tem como exceção o estado de Santa Catarina que se encontra na faixa intermediária.

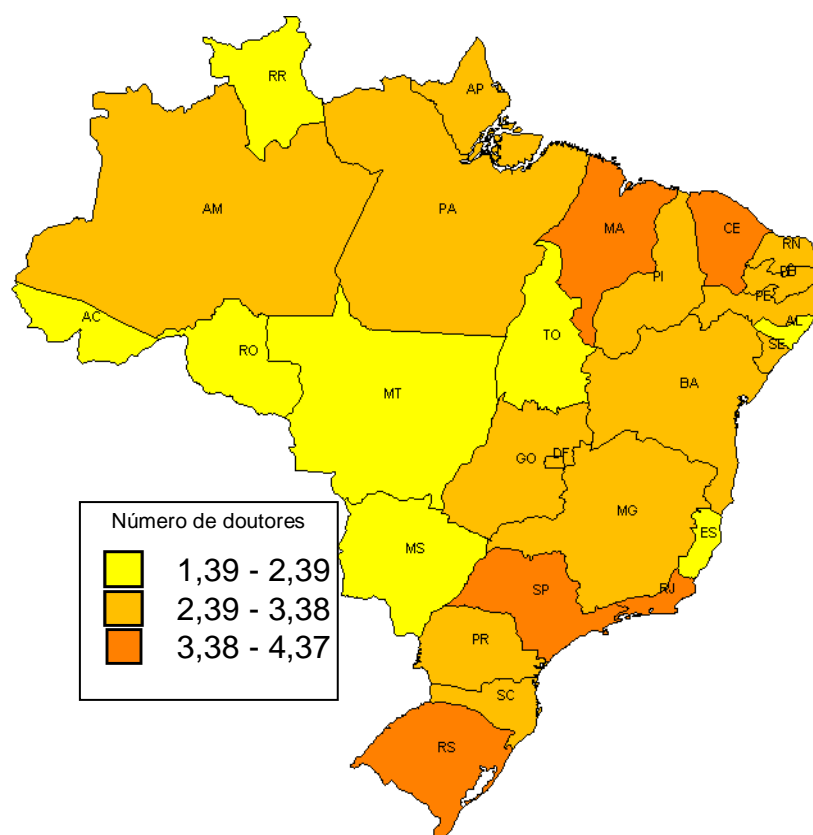
O mapa temático, como ferramenta de visualização de dados, permite um maior entendimento e/ou uma nova leitura dos dados. Já no histograma do Gráfico 6.3.11, observado anteriormente, foi possível identificar uma distribuição homogênea com estados de todas as regiões figurando entre os mais produtivos do país.



**Figura 7.1.1** - Artigos completos publicados em periódicos especializados de circulação nacional

Vale ressaltar que o estado de Roraima supera, em termos da produção científica relativa ao número de doutores, estados economicamente e teoricamente mais desenvolvidos como, Rio de Janeiro e Santa Catarina e Bahia. Já o estado de Alagoas é o único do Nordeste que aparece na faixa de menor produtividade.

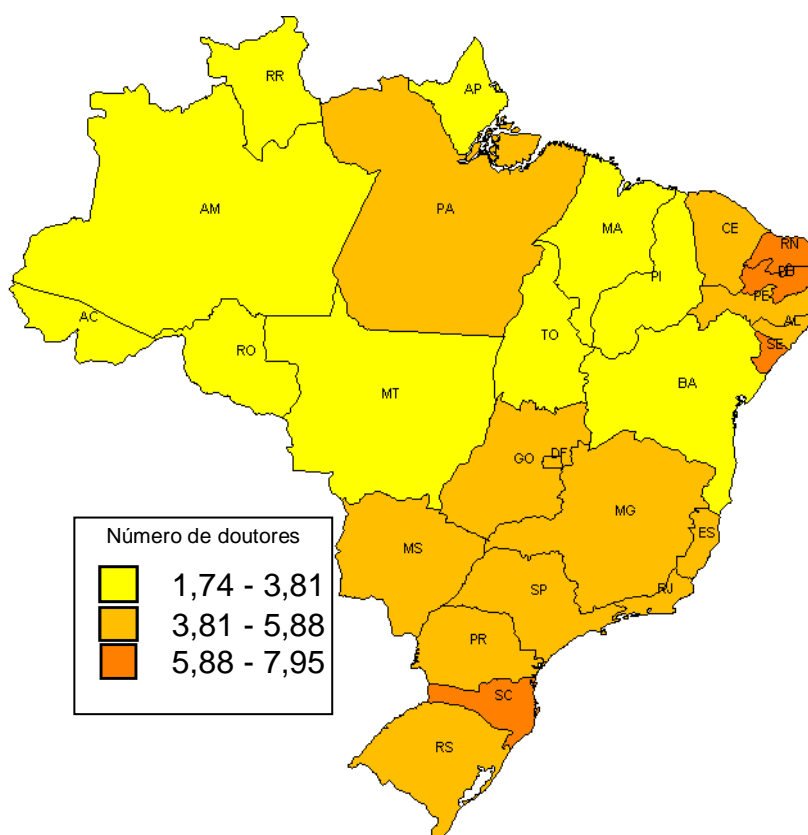
A publicação de artigos em periódicos Internacionais mantém suas faixas de produtividade próxima, com 0,99 pontos que representa a distância entre o valor mínimo e máximo dentro da faixa, demonstrado no mapa que mesmo os menos produtivos estão muito próximos dos com maior produtividade.



**Figura 7.1.2** - Artigos completos publicados em periódicos especializados de circulação internacional

O diferencial com a produção de artigos em periódicos nacionais está no número de estados que estão na faixa mais produtiva, apenas cinco aparecem sendo que quatro deles se mantiveram nessa faixa. Os estados do Acre e Alagoas continuam na faixa de menor produtividade e Roraima que mereceu destaque está entre os mais produtivos para artigos nacionais, passou para a menor faixa em artigos internacionais.

No mapa temático relacionado a trabalhos completos em anais de evento pode ser observada uma variação de faixa de aproximadamente 2,07 pontos, mais que o dobro dos mapas anteriores. Destacam-se em produtividade os estados de Santa Catarina (região Sul), e o Rio Grande do Norte, Paraíba e Sergipe (região Nordeste), que aparecem na faixa de maior produtividade. Na região Norte o Pará se destaca por ser o único estado a se enquadrar, pelo menos, na faixa intermediária de produtividade.

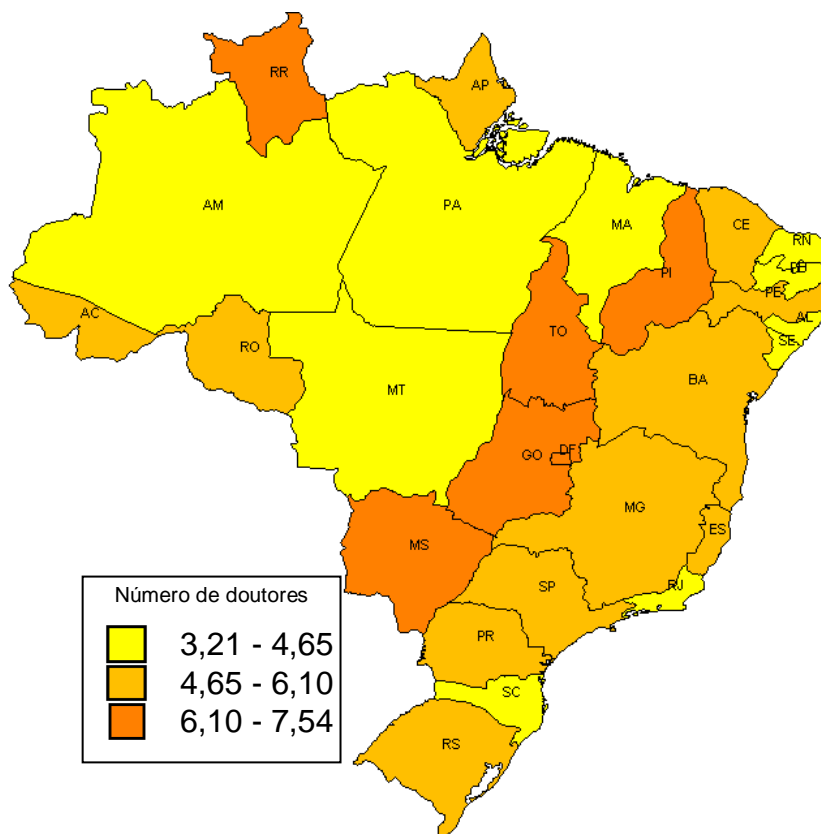


**Figura 7.1.3** - Trabalhos completos publicados em anais de eventos

Os mapas a seguir (Figura 7.1.4 e Figura 7.1.5) representam dados da produção de livros e capítulos de livros. O primeiro tem uma variação de faixa insignificante de 0,09 exibindo os menores índices onde o maior apresenta o valor de 0,47, menos de meio livro por doutor. Esse mapa só explica que todos os estados têm baixa produtividade para esse indicador. Destaque para os que aparecem na faixa mais produtiva dois do Sul e um do Norte e outro do Nordeste além do Distrito Federal.



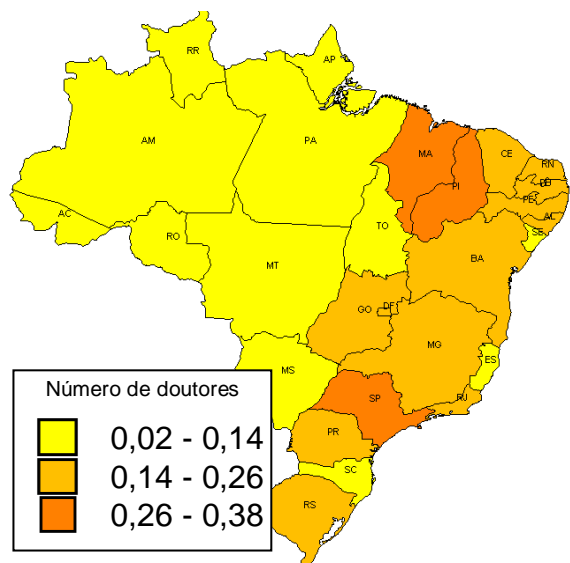




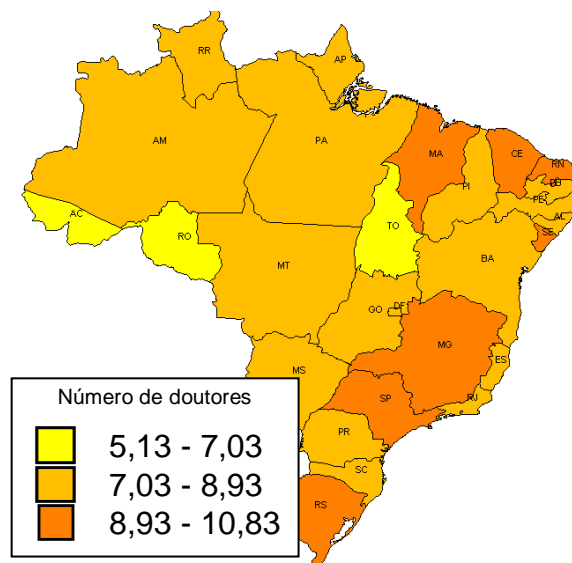
**Figura 7.1.6** - Outras publicações bibliográficas

Os resumos publicados em periódicos especializados mostrados na Figura 7.1.7 também possuem baixo índice de publicações no Brasil. Apresenta os estados de São Paulo, Maranhão e Piauí na maior faixa produtiva, mas a variação entre as faixas é muito pequena. Percebe-se que a produtividade na primeira faixa é tão pequena que não alcança 0,2 de produção por doutor e abrange o maior contingente de estados. Talvez seja um indicador que deva ser repensado enquanto significativo para representar produção ou ser fonte de estudos mais específicos já que o maior índice não chega a meio resumo por doutor.

Apresentando os maiores índices de produtividade, começando em cinco e fechando com até dez publicações por doutor, o mapeamento dos dados de resumos em anais (Figura 7.1.8), tem uma variação entre faixas de 1,90 pontos. Outra grata surpresa foi encontrar apenas três estados (Acre, Rondônia e Tocantins) na menor faixa de produtividade sendo que essa vai de cinco a sete publicações por doutor. Fechando essas análises com este último mapa, percebeu-se que os estados do Acre, Alagoas, Amazonas e Mato Grosso foram os únicos que não alcançaram a faixa de maior produtividade em nenhum dos indicadores.



**Figura 7.1.7** - Resumos de trabalhos publicados em periódicos especializados



**Figura 7.1.8** - Resumos de trabalhos publicados em anais de eventos

A análise dos mapas temáticos demonstra um cenário muito uniforme entre os estados. Existem destaques em um e outro mapa, mas a percepção geral é que todos são muito próximos devido à pequena variação entre as faixas de valores. Quando os valores absolutos são transformados em índices, para cada estado, as altas taxas de produção, como as apresentadas por São Paulo, convertidas ficam muito próximas das demais e até mesmo inferiores em alguns casos.

Mesmo que esses ainda se mantenham entre os mais produtivos, como no caso do estado de São Paulo, os índices apresentados estão próximos de outros estados e até são superados. Os estados com poucos doutores demonstram que produzem tanto quanto os que possuem muitos e em alguns casos até mais.

Para concluir, os mapas desenvolvidos nesta pesquisa apresentam praticamente as mesmas características comuns entre eles com poucos grupos e grande similaridade entre os estados. Esses mapas demonstram um equilíbrio entre os estados, talvez, demonstrando que quando números absolutos são convertidos em relativos, apresentam os dados de forma mais clara sem a “fascinação” gerada por números grandiosos.

## 8 ANÁLISE DE DADOS MULTIVARIADOS

Na seção anterior foram utilizados métodos da estatística univariada, mas existe um momento da análise de dados que se faz necessário conhecer as variáveis de forma conjunta. Esse método é a estatística multivariada. A denominação estatística ou análise multivariada corresponde a um grande número de métodos e técnicas que utilizam simultaneamente todas as variáveis na interpretação teórica do conjunto de dados obtidos. (MOITA NETO, 2004).

O desenvolvimento tecnológico oriundo das descobertas científicas tem alavancado o próprio desenvolvimento científico, ampliando em várias ordens de grandeza a capacidade de obter informações de acontecimentos e fenômenos que estão sendo analisados (MOITA NETO, 2004). O uso de softwares estatísticos vem, dentre outras coisas, facilitando o trabalho de pesquisadores que necessitam entender como se relacionam as variáveis e qual a importância destas dentro do estudo.

O estudo multivariado apresenta mecanismos que ajudam a interpretar o conjunto de dados permitindo ao pesquisador traduzir as informações ocultas nessa massa e por isso não são percebidas.

Nas próximas seções, será apresentada a aplicabilidade desses métodos, dividida em itens, servindo de suporte para o desenvolvimento das técnicas apresentadas. No primeiro item, será discutida a análise das componentes principais. No segundo, será apresentado o método de faces de Chernoff. E no último item, será discutida a aplicação da lógica difusa. Para todos os itens foram abordados sua fundamentação e suas aplicações no estudo da produção científica nacional.

## 9 ANÁLISE DE COMPONENTES PRINCIPAIS

A análise de componentes principais consiste em reescrever as variáveis originais em novas variáveis denominadas componentes principais, através de uma transformação de coordenadas (MOITA NETO, 2004). Para uma compressão das variáveis essa técnica projeta o máximo de informações na menor quantidade de eixos possíveis não correlacionados. A análise de componentes principais é uma técnica estatística poderosa que pode ser utilizada para redução do número de variáveis e para fornecer uma visão estatisticamente privilegiada do conjunto de dados. (MOITA NETO, 2004).

Para padronizar os valores dos dados foi utilizado o método estatístico percentil (“z” *percentis*) que refere-se à posição ocupada por determinada observação no interior de uma distribuição. Tal procedimento é calculado utilizando a média, desvio padrão pela seguinte fórmula:

$$z = \frac{X - \mu}{\sigma} \rightarrow \text{Escore-z} = \frac{\text{Valor observado} - \text{Valor da mediana de referência}}{\text{Desvio padrão dos dados de referência}}$$

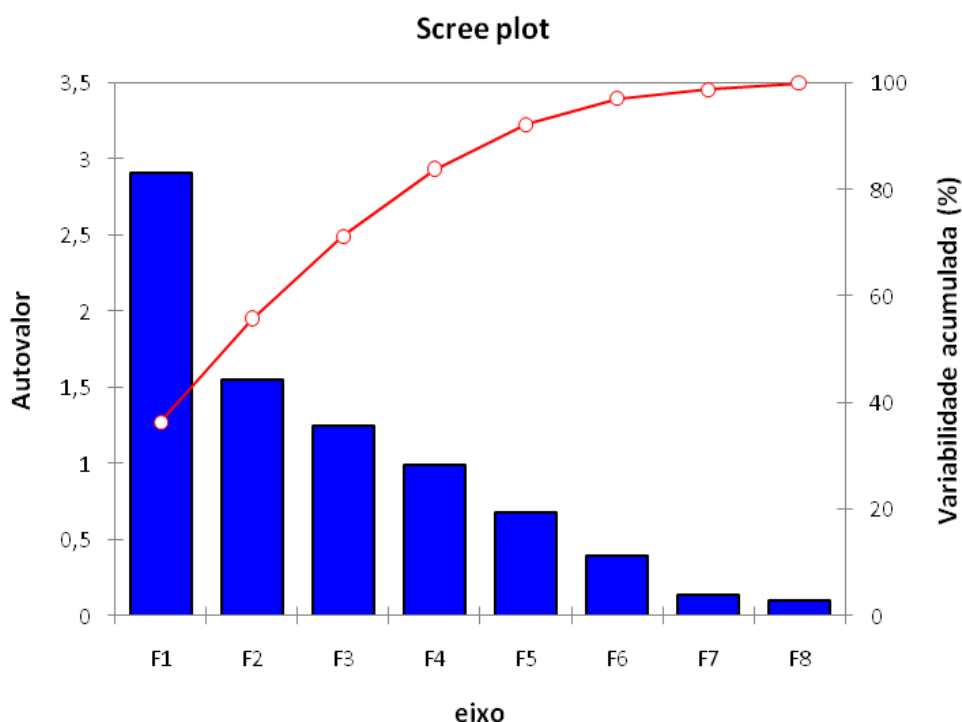
A partir das combinações lineares das variáveis originais, com pesos escolhidos, é maximizada a variância com a primeira componente (F1) representando a direção de máxima variância no espaço multidimensional para os dados representados e a segunda componente (F2), ortogonal à primeira, descreve o máximo da informação ainda não descrita por (F1).

**Tabela 9.1- Contribuições das variáveis (%) em relação as componentes F1 e F2**

Variáveis	F1	F2
Artigos nacional	12,788	28,103
Artigos internacional	25,244	4,981
Trabalhos completos anais	10,593	0,565
Livros	4,955	0,496
Capítulos de livros	8,932	12,180
Outras publicações	0,155	46,491
Resumos Periódicos	19,175	3,542
Resumo Anais	18,158	3,643

A análise de componentes principais fornece as ferramentas adequadas para identificar as variáveis mais importantes no espaço das componentes principais (MOITA NETO, 2004). Na análise dos dados da produção científica cada indicador de produção científica pode ser imaginado geometricamente como um ponto localizado num sistema de oito dimensões (uma para cada variável). No Scree plot<sup>4</sup> (Gráfico 9.1) apresentam-se a variabilidade acumulada e os pesos das componentes.

Este método permite a redução da dimensionalidade dos pontos representativos das amostras pois, embora a informação estatística presente nas n-variáveis originais seja a mesma dos n-componentes principais, é comum obter em apenas 2 ou 3 das primeiras componentes principais mais que 90% desta informação. O gráfico da componente principal 1 versus a componente principal 2 fornece uma janela privilegiada (estatisticamente) para observação dos pontos no espaço n-dimensional (MOITA NETO; MOITA, 1998, 468)

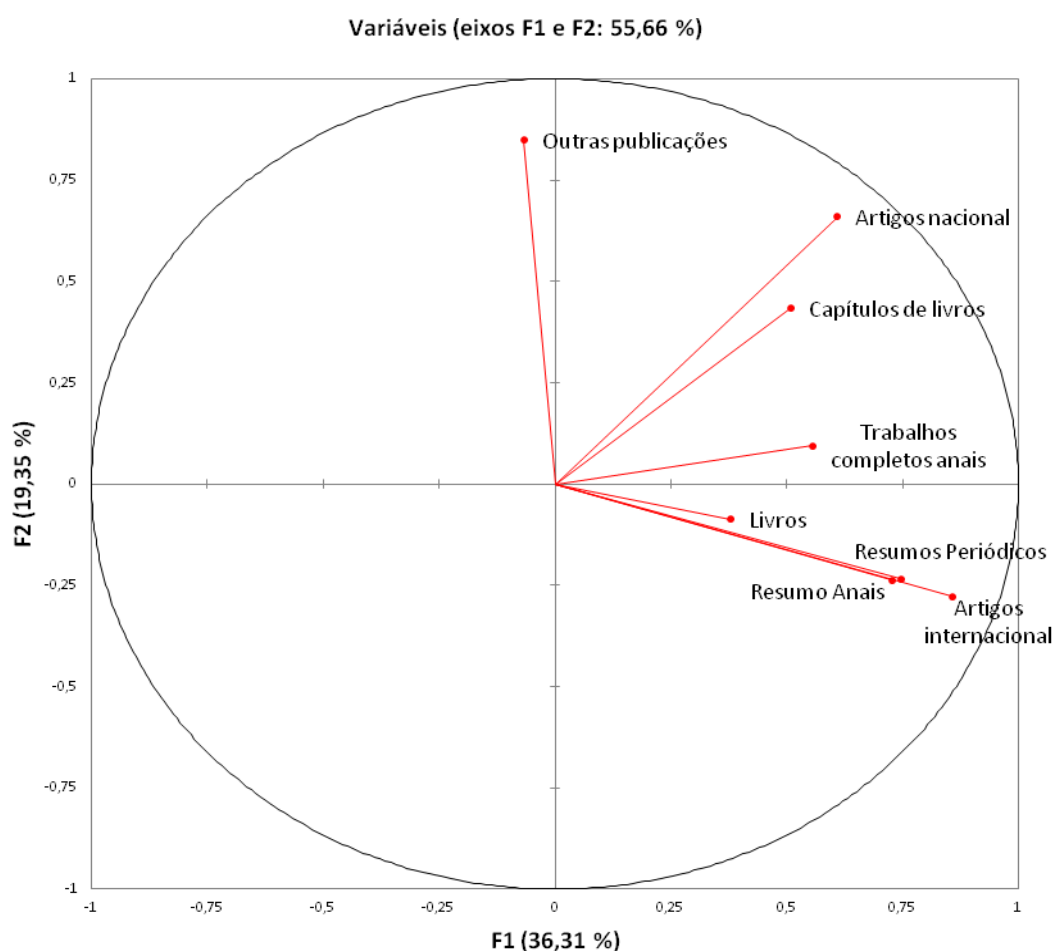


**Gráfico 9.1** - Scree plot das componentes principais

Através dos gráficos é possível perceber padrões entre as variáveis. O Gráfico 9.2 indica que as duas primeiras componentes juntas explicam mais de 50% de toda a informação. A primeira componente (36,31%) determina que todas as variáveis, exceto outras publicações, têm pesos positivos. A aproximação do ponto

<sup>4</sup> Este gráfico permite identificar as componentes que formaram os eixos mais significativos dentro da variabilidade dos dados e sua distribuição entre os eixos da ordenação.

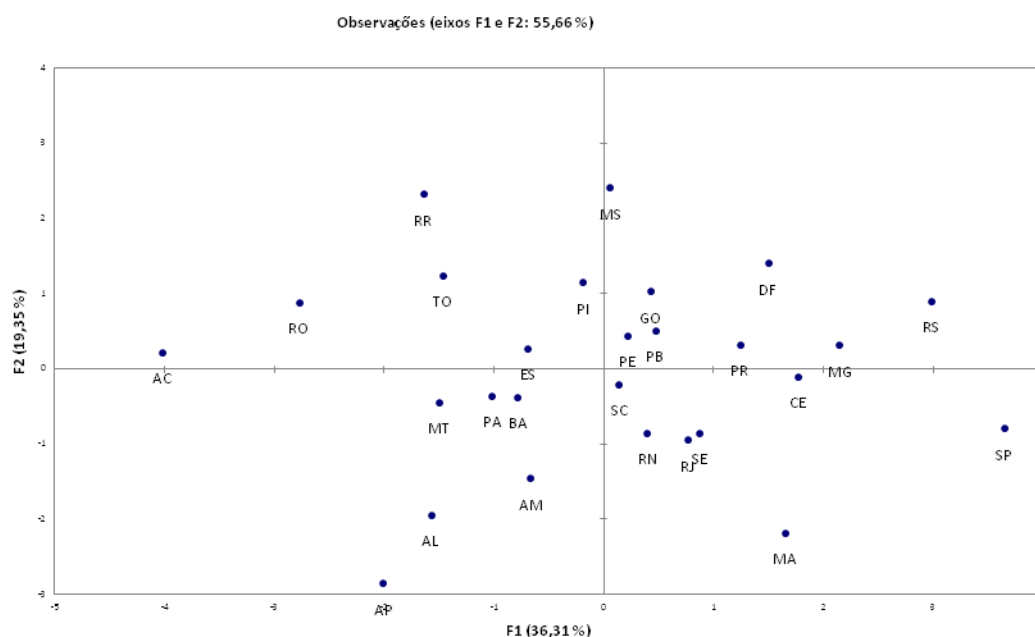
ao círculo indica que a variável está mais bem representada no plano que nesse caso são artigo nacional e internacional. A proximidade dos pontos entre si indica que estas variáveis têm uma alta correlação positiva. Nessa análise fica evidente que entre as variáveis existem relações interessantes, como a de que capítulos de livros aproxima-se de artigos nacionais e livros aproxima-se de artigos internacionais. Da mesma forma, a localização de variáveis no quadrante oposto indica correlações negativas. Isso implica que para as variáveis artigos nacional e artigos internacional a relação é inversa.



**Gráfico 9.2** – ACP das variáveis

Analisando a partir do posicionamento de cada estado observa-se no primeiro e quarto quadrantes (valores positivos em F1) os estados mais produtivos; no segundo e terceiro quadrantes (negativos em F1) os menos produtivos. Já a

segunda componente indica certa vocação dos estados em desenvolver mais determinado tipo de produção. Fica claro ao observar que São Paulo segue na direção de artigos internacional. A partir dos quadrantes é possível dizer que os estados que aparecem no primeiro e quarto são os mais bem posicionados quanto à produtividade por se aproximarem mais das variáveis mais representativas e possuir valores positivos.



**Gráfico 9.3** – ACP das observações

Depois de identificar a relação das variáveis no universo das componentes principais e as aproximações e distâncias dos estados em relação a esses indicadores de produção é perceptível a importância desse método para a análise multivariada de dados indicados por Moita Neto explicando que

As variáveis originais têm a mesma importância estatística, enquanto que as componentes principais têm importância estatística decrescente. Ou seja, as primeiras componentes principais são tão mais importantes que podemos até desprezar as demais. Destas características podemos compreender como a análise de componentes principais: a) podem ser analisadas separadamente devido à ortogonalidade, servindo para interpretar o peso das variáveis originais na combinação das componentes principais mais importantes b) podem servir para visualizar o conjunto da amostra apenas pelo gráfico das duas primeiras componentes principais, que detêm maior parte da informação estatística. (MOITA NETO, 2004).

Com esses dados é possível estabelecer um critério de seleção de variáveis para aplicação de outros métodos complementares. A força apresentada por indicadores como artigo nacional, internacional, capítulo de livros e resumos (em periódicos e anais) expressa que essas variáveis são realmente expressivas para indicar produtividade e perfis dos estados. Já o resultado encontrado para outras publicações mal posicionado só demonstra que esse indicador, que parece um refugio de tudo que é produzido sem uma classificação própria, pode ser descartado numa posterior análise.

Seguindo o esforço de demonstrar as facilidades do uso de tecnologias para desenvolver os processos matemáticos e estatísticos nesse estudo, esta a disposição no apêndice 1 o relatório de análise dos dados gerados pelo *software* XLStat.






## 10 TÉCNICAS ICONOGRÁFICAS

As técnicas iconográficas utilizam uma entidade gráfica dotada de propriedades geométricas (forma, tamanho, posição e direção) e características de aparência (cor, textura, transparência) que são modificadas de acordo com o valor do atributo do dado que representam. Por exemplo, a largura e a altura de uma caixa poderiam ser determinadas pela pontuação de um aluno nos exames de um curso, enquanto a cor da caixa pode indicar o sexo do aluno (WARD, 2008).

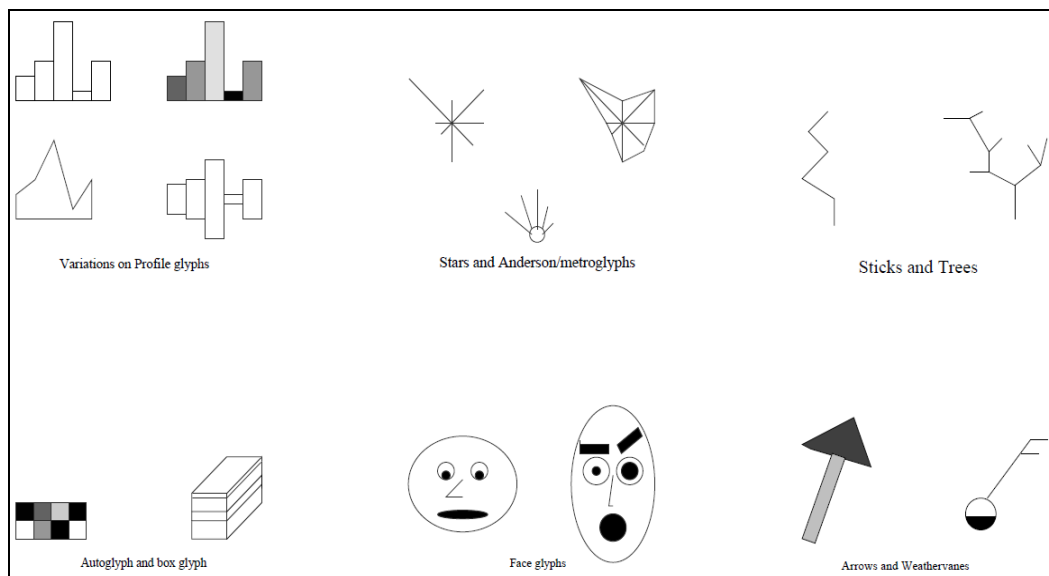
Exemplo:

Nota 1 – altura | Nota 2 – largura | cor – sexo = azul – homem, vermelho - mulher

Aluno	Representação
Paulo	
Rita	
Marcos	

Podemos exemplificar como as técnicas iconográficas são discutidas das seguintes formas técnicas: (a) Faces de Chernoff; (b) Star Glyphs e (c) Figura de Arestas. A técnica de Star Glyphs pode representar um número maior de atributos quando comparada a Faces de Chernoff, mas ambas apresentam restrições quanto ao número de variáveis representáveis.

Glifo, do grego *glyphé* (gravura), é a denominação utilizada para designar pictograma gravado em pedra. Essa denominação também é utilizada para representar desenhos simplificados utilizados para apresentar graficamente dados multidimensionais complexos. Como exemplos de glifos existentes temos uma lista indicada por Ward (2008) que servem de ilustração na Figura 10.1.



**Figura 10.1** - Exemplos de glifos multivariada segundo Ward (2008)

No entanto, glifos têm suas limitações. Eles geralmente são restritos em termos de como exatamente eles podem transmitir dados devido ao seu tamanho e os limites do sistema de percepção visual humano para medir diferentes atributos gráficos. Há também restrições sobre o número de registros de dados que podem ser efetivamente visualizado com glifos; tamanho do conjunto de dados excessivo pode resultar em oclusão significativa ou a necessidade de reduzir o tamanho de cada glifo, os quais fazem a detecção de padrões difícil, se não impossível. Assim glifos são principalmente adequados para a análise qualitativa dos conjuntos de dados de tamanho modesto (WARD, 2008).

A utilização dessa técnica apresenta-se muito pertinente para analisar a produção científica dos estados brasileiros a fim de perceber suas semelhanças e diferenças. O conjunto de variáveis é numérico e pequeno. Duas dimensões são o bastante para uma interpretação qualitativa desses dados. E dentre os modelos de glifo, as faces de Chernoff demonstram uma maior relação com a sensibilidade de percepção humana e mais própria para esse estudo, uma vez que se objetiva, também, apresentar e indicar tais técnicas para aplicação na Ciência da Informação.

Utiliza-se a denominação glifo para um objeto geométrico que, representando uma entidade ou elemento de amostragem, tem a forma e outros atributos visuais determinados pelos valores dos atributos da entidade. Um exemplo clássico de utilização de glifos é o conjunto de faces de Chernoff (FREITAS, 2001)

Faces de Chernoff é um método utilizado para representar dados multivariados utilizando faces cujas características são determinadas pelas coordenadas do ponto. Para fazer uma análise descritiva de dados uma determinada coordenada corresponde ao comprimento de um dos elementos da face como nariz, olho, boca, etc. Esse processo facilita tanto a percepção de similaridades e/ou irregularidades presentes nos dados, como a classificação ou agrupamento indivíduos. Foi Herman Chernoff, em 1971, quem criou essa forma de representar dados n-dimensionais.

Um método de representação gráfica de pontos no espaço k-dimensional é apresentado para apreciação. O objetivo é representar dados multivariados, sujeitos a fortes, mas possivelmente relações complexas, de tal forma que um investigador pode rapidamente compreender as informações pertinentes e, em seguida, aplicar a análise estatística adequada. (CHERNOFF,1973)<sup>5</sup>

Observando que o ser humano é sensível ao reconhecimento de expressões faciais, Chernoff (1973) criou uma representação gráfica cujos atributos de dados são mapeados para as propriedades de uma face. As faces de Chernoff podem ser utilizadas para identificar agrupamentos, por meio das semelhanças entre as faces, detectar padrões nos dados e identificar semelhanças e diferenças pela verificação de cada características das faces.

Nos estudos de Chernoff (1973), existem dois exemplos de aplicação desta técnica. Primeiro, o especialista do domínio estava interessado em encontrar agrupamentos nos dados, e segundo, outro analista desejava observar o comportamento dos dados em relação ao tempo por meio da seqüência de faces apresentadas.

Os valores dos atributos são responsáveis pelas alterações das propriedades das faces, como posição dos olhos, tamanho do nariz, abertura da boca, dentre outras configurações permitidas. Desse modo, é conveniente que os valores dos atributos sejam numéricos, para que seja possível realizar o mapeamento para as propriedades da face.

---

<sup>5</sup> A graphical method of representing points in k-dimensional space is presented for consideration. The object is to represent multivariate data, subject to strong but possibly complex relationships, in such a way that an investigator can quickly comprehend relevant information and then apply appropriate statistical analysis.

Dentre trabalhos que apresentam a aplicação das faces de Chernoff estão: os que representam dados de medições arqueológicas (dados quantitativos discretos), os que utilizam dados financeiros (dados quantitativos contínuos) e os que apresentam mapeamento de dados populacionais (dados quantitativos contínuos) para as faces de Chernoff.

## 10.1 A APLICAÇÃO DAS FACES DE CHERNOFF AOS DADOS

Segundo Chernoff (1973)<sup>6</sup> o método consiste em representar um ponto no espaço k-dimensional por um desenho de um rosto, cujas características são determinadas pela posição do ponto. Para essa análise foi utilizado o software Faces de Chernoff desenvolvido pelo professor Jorge Luiz Perrone<sup>7</sup>. O software foi muito útil e já apresentou, ao analisar os dados, faces com a formação de grupos.

A única limitação é a possibilidade de utilizar apenas seis variáveis, como visto anteriormente essa já era uma limitação conhecida no uso dessa técnica. o que levou a necessidade de descartar duas variáveis. Para solução foram utilizados os resultados propostos pela Análise de Componentes Principais os quais indicavam as variáveis que estavam mais bem representadas e quais apresentavam inter-relações mais fortes.

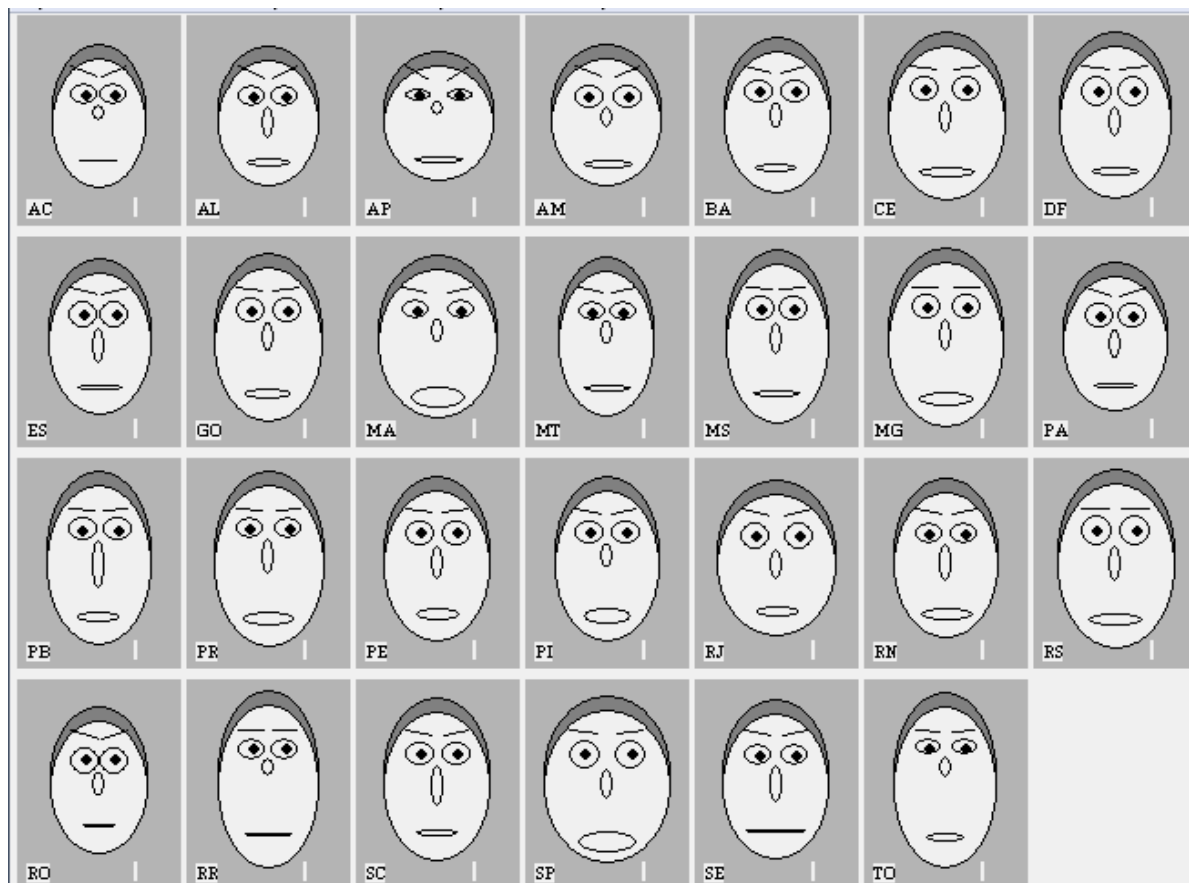
Para aplicar as Faces de Chernoff aos dados foi feita a seguinte relação entre os elementos da face e as variáveis da seguinte forma:

Altura do rosto	→	Artigos em Periódicos Nacionais
Largura do Rosto	→	Artigos em Periódicos Internacional
Nariz	→	Trabalhos Completos publicados em Anais de Eventos
Olhos	→	Capítulos de Livros
Abertura da boca	→	Resumos em Periódicos Especializados
Tamanho da boca	→	Resumos em Anais

**Quadro 10.1.1** – Definição das variáveis da face

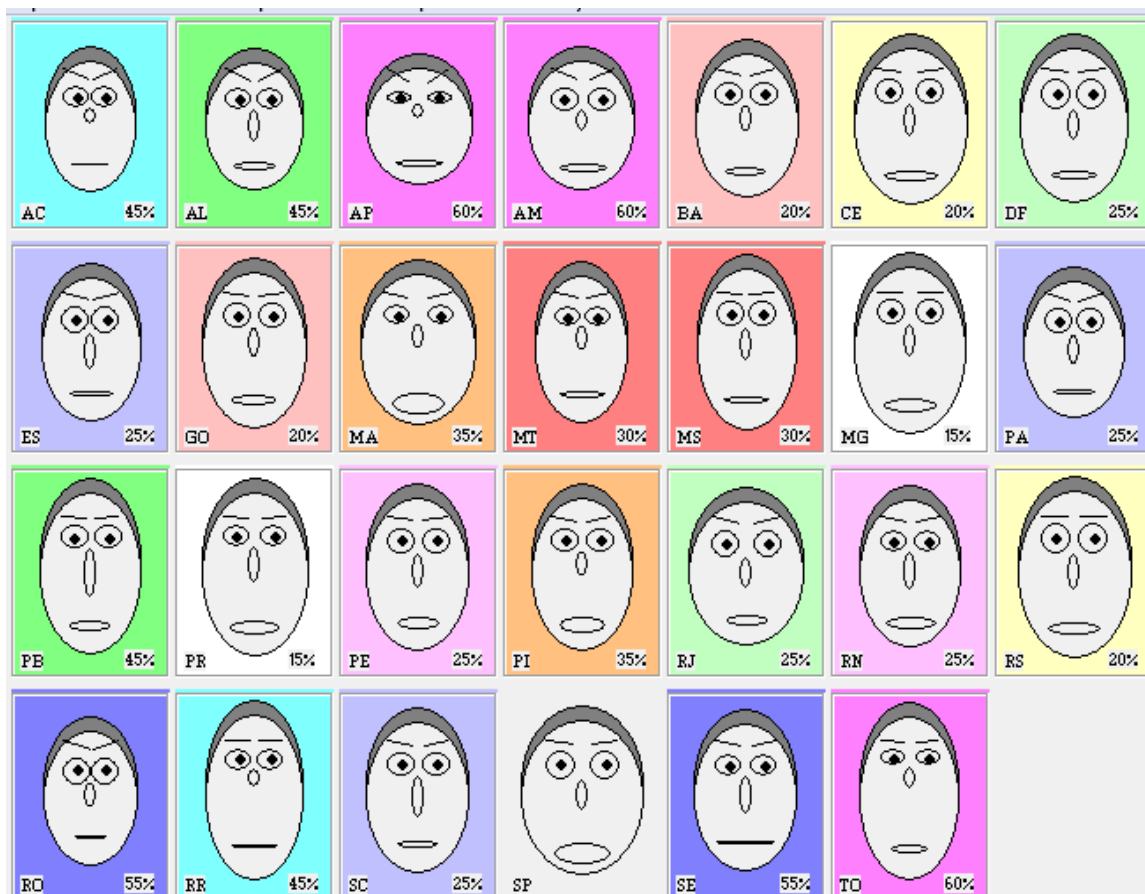
<sup>6</sup> the method consists of representing a point in k-dimensional space by a drawing of a face whose characteristics are determined by the position of the point.

<sup>7</sup> Mestre em Modelagem Computacional e professor titular da Universidade Católica do Salvador.



**Figura 10.1.1** - Faces de Chernoff da produção científica dos estados brasileiros

Observando as faces é possível identificar as similaridades. Se for observado um elemento da face de cada vez é fácil perceber que existe em pelo menos um similar em outra face. Altura do rosto, nariz e abertura e tamanho da boca aparecem praticamente iguais em várias faces vistas individualmente. Feita a análise a partir de um elemento conhecido como a boca, por exemplo, que representa os resumos (abertura e tamanho) é fácil perceber que muitos estados ficam de boca fechada literalmente, ou seja, não são grandes produtores e são similares nesse quesito. A figura 10.1.1, a aparência que seus dados dão a face utilizando a representação de cores e valores percentuais.



**Figura 10.1.2** - Faces de Chernoff da produção científica dos estados brasileiros com identificação de grupos

As faces de Chernoff realçam ainda mais a visualização dessa similaridade, observando o perfil dos estados é fácil perceber a formação de poucos grupos e que esses são formados por membros de regiões geográficas diferentes. Esse agrupamento é muito mais característico do que a aproximação/vizinhança geográfica, uma vez que, dentro das regiões territoriais existem perfis muito distintos em alguns casos.

Princípios da Gestalt indicam que temos uma preferência por formas simples, e que seres humanos tentam sempre encontrar alguma simetria. Avaliação da eficácia de glifos para a apresentação de informação e análise pode ser realizada em um número de maneiras diferentes. Um grande número de estudos de avaliação sobre glyphs e outras formas de análise multivariada de dados têm sido realizados ao longo dos anos. Alguns destes têm sido bastante ad hoc, ou com base apenas nas opiniões e observações dos autores, enquanto outros têm envolvido detalhada e cuidadosamente orquestrada estudos de usuários (WARD, 2008).

Ainda segundo Ward (2008), várias avaliações de faces Chernoff foram relatados desde a sua criação. Diversificadas análises de sua aplicação se concentraram no estudo da eficácia e da pré-atenção, por parte dos seres humanos, de diferentes características faciais. A quantidade do tempo necessário para interpretar os dados foi medida e analisada obtendo excelentes resultados. A conclusão foi que os estudos de Chernoff podem ser comparados aos outros tipos de glifos multivariados e obter algumas vantagens

## 11 A LÓGICA DIFUSA

A lógica difusa encontra-se entre as técnicas mais recentes de Inteligência Artificial, também conhecida como conjuntos difusos. Este termo, a princípio, nos convida a pensar em algo confuso (nebuloso), porém a lógica difusa consiste em aproximar a decisão computacional da decisão humana, tornando as máquinas mais capacitadas a seu trabalho.

Uma das principais potencialidades da lógica difusa, quando comparada com outros esquemas que tratam com dados imprecisos é que suas bases de conhecimento, as quais estão no formato de regras de produção, são fáceis de examinar e entender. Este formato de regra também torna fácil a manutenção e a atualização da base de conhecimento.

O conceito de conjunto difuso foi introduzido em 1965 por Lotfi A. Zadeh (Universidade da Califórnia, Berkeley). Zadeh é reconhecido como grande colaborador do Controle Moderno. Em meados da década de 60, Zadeh observou que os recursos tecnológicos disponíveis eram incapazes de automatizar as atividades relacionadas a problemas de natureza industrial, biológica ou química, que compreendessem situações ambíguas, não passíveis de processamento através da lógica computacional fundamentada na lógica booleana.

A antecedente desses estudos é a tendência para a formulação de alguns problemas de controle no contexto de sistemas de estado finito, isto é, sistemas cujo espaço de estado consiste em um número finito de pontos em vez de um contínuo. A explicação era que tais sistemas são geralmente mais facilmente passíveis de análise e simulação por computadores digitais do que os de estado contínuo, e particularmente bem adequados para a aplicação das técnicas de programação dinâmica (ZADEH, 1963).

Em 1965, Zadeh introduziu os fundamentos da lógica difusa<sup>8</sup> (lógica difusa) que se baseia em regras que toleram a imprecisão e é utilizada dessa para a

---

<sup>8</sup> Clearly, the "class of all real numbers which are much greater than 1," or "the class of beautiful women," or "the class of tall men," do not constitute classes or sets in the usual mathematical sense of these terms. Yet, the fact remains that such imprecisely defined "classes" play an important role in human thinking, particularly in the domains of pattern recognition, communication of information (ZADEH, 1965).



solução de problemas complexos. Diferente da lógica clássica, em que um objeto possui apenas duas possibilidades (ser ou não ser elemento do conjunto), a lógica difusa, com a teoria dos conjuntos nebulosos, apresenta para o mesmo objeto vários graus de pertinência ao conjunto. Vulgarmente chamada de lógica nebulosa, essa lógica permite o tratamento de expressões que envolvem grandezas descritas de forma não exata.

O modelo para a lógica difusa é a mente humana. Ao lidar com problemas do mundo real existe muito a ser ganho com a exploração da tolerância à imprecisão. E essa é a principal motivação do desenvolvimento da lógica difusa (ZADEH, 1996).

Ao estudar as aplicações da lógica difusa torna-se mais fácil perceber e entender sua utilização na otimização de problemas complexos e sua aplicação no entendimento do complexo e diversificado perfil da produção científica do Brasil. A principal contribuição da lógica difusa é uma metodologia para computação com palavras (ZADEH, 1996).

Em seu sentido tradicional, a lógica difusa envolve computação (na sua maior parte), e a manipulação de números e símbolos. Em contraste, os seres humanos empregam principalmente palavras durante um raciocínio, chegando a conclusões expressas como palavras a partir de premissas expressas em uma linguagem natural ou que tenham a forma de percepção mental. Como o usado por seres humanos, as palavras têm valor referencial na lógica difusa (ZADEH, 1996).

Os seres humanos, que muitas vezes dependem de expressões imprecisas como "geralmente", "caro", ou "longe", a compreensão de um computador é limitado a um preto-branco, modo tudo-ou-nada, ou verdadeiro ou falso. Neste contexto, Lotfi Zadeh enfatiza o fato de que facilmente nos deixamos ser arrastados por um desejo de atingir a precisão mais alta possível, sem prestar atenção ao caráter impreciso da realidade.

Um passeio pela lógica ocidental nos revela que o pensar binário (certo ou errado, verdadeiro ou falso, etc) não atende a todas as questões uma vez que não permite a possibilidade da resposta ser ao mesmo tempo parcialmente verdadeira e parcialmente falsa. "Na verdade, entre a certeza de ser e a certeza de não ser, existem infinitos graus de incerteza. Esta imperfeição intrínseca à informação

representada numa linguagem natural tem sido tratada matematicamente no passado com o uso da teoria das probabilidades” (UNIVERSIDADE ESTADUAL DE MARINGÁ, 2006).

Essa lógica trata com valores “verdade” das afirmações, classificando-as como verdadeiras ou falsas, mas em muitos casos não é possível através dessa simples classificação de verdadeiro ou falso / sim ou não se fazer uma afirmação. Por exemplo, determinar se um ambiente está frio ou quente, se uma pessoa é alta ou baixa ou se uma nota atribuída a um aluno é boa ou ruim. Um simples sim ou não na maioria das vezes não responderá de forma completa a questão sem a utilização de parâmetros.

Em caracterizações linguísticas de variáveis e suas dependências, as palavras servem como os valores das variáveis e desempenham o papel de restrições da lógica difusa. Nesta perspectiva, o uso das palavras pode ser visto como uma forma de granulação, que por sua vez pode ser considerada como uma forma de quantização difusa (ZADEH, 1996).

A lógica difusa contradiz esse princípio do determinismo binário e “o conceito de dualidade, estabelecendo que algo pode e deve coexistir com o seu oposto, fazendo com que a lógica difusa pareça natural, até mesmo inevitável ” (UNIVERSIDADE ESTADUAL DE MARINGÁ, 2006). Sua forma de análise baseada em pesos e parâmetros aplicados a tais dados foi de muita utilidade na apresentação da realidade da produção científica brasileira.

Dentre as características que podem ser destacadas temos: a lógica difusa que tem como base termos e não números, ou seja, os valores verdadeiros são expressos linguisticamente (quente, muito frio, verdade, longe, perto, rápido, vagaroso, médio, etc), o uso das probabilidades linguísticas (provável, improvável) que são interpretados como números gerados a partir da lógica difusa e manipulados pela sua aritmética utilizando valores apenas com limites, possui vários modificadores de predicado (muito, mais ou menos, pouco, bastante, médio, etc), e também um amplo conjunto de quantificadores, como por exemplo: poucos, vários, em torno de, usualmente.

## 11.1 A APLICAÇÃO DA LÓGICA DIFUSA AOS DADOS DA PRODUÇÃO CIENTÍFICA

Para analisar os dados da produção científica nacional foram utilizados os indicadores mais relevantes não só pelo apelo aspecto significativo que estes apresentam no contexto de divulgação da informação científica, mas também pelo fato que nesse momento da pesquisa, já existem subsídios suficientes para uma escolha dos indicadores mais significativos para comporem as variáveis a serem analisadas.

Anteriormente, durante a análise dos mapas temáticos foi possível perceber que muitos indicadores apresentaram uma produção muito pequena a partir dos valores *per capita* tornando os dados similares entre os estados. Posteriormente, a análise das componentes principais apresentou quais indicadores eram representativos para uma melhor análise de seus dados.

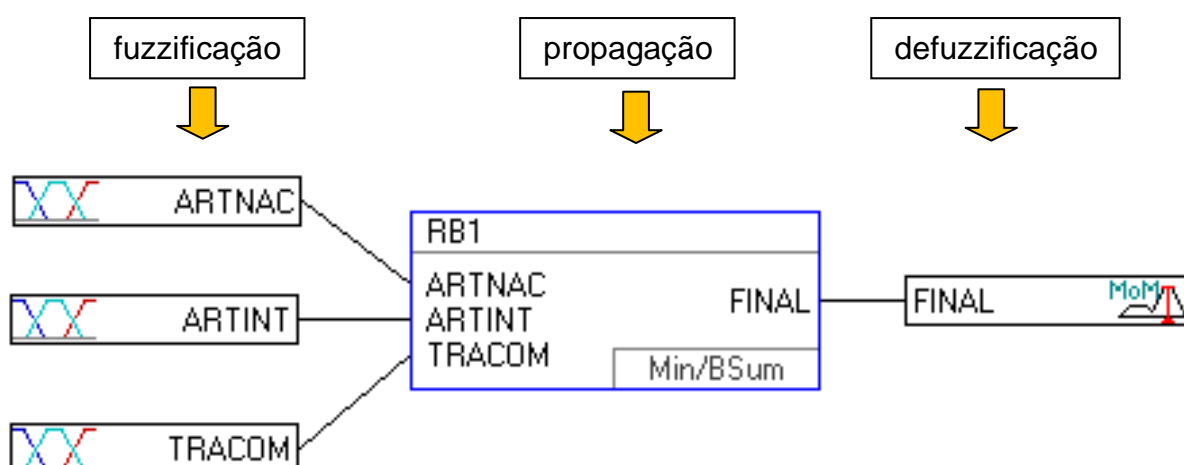
Com base no exposto para a utilização do método difuso no estudo da produção científica foram escolhidas apenas três variáveis: produção de artigos em periódicos internacional, de artigos em periódico nacional e trabalhos completos em anais, as quais foram representados respectivamente por ARTINT, ARTNAC e TRACOM.

A aplicação da lógica difusa aos dados aconteceu em 03 (três) etapas. No primeiro, a fuzzificação que faz a conversão do valor fixo de entrada em graus de cada qualificador, ou seja, nesta etapa definem-se as variáveis linguísticas de forma subjetiva bem como as funções pertinência. Dentro dessa etapa acontece a análise do problema, a definição das variáveis difusas, a definição das funções de pertinência e a criação das regiões.

Na segunda etapa, da propagação, é o momento em que se aplicam as regras difusas nos qualificadores. Essa etapa engloba a definição das proposições, a análise das regras difusas, a criação da região resultante. E no último estágio, da defuzzificação, é o momento que os graus resultantes são convertidos para valores fixos, ou seja, as variáveis difusas são convertidas em valores numéricos ou aceitáveis pelo sistema, etapa em que diversas técnicas de defuzzificação podem ser usadas.

A fuzzificação e defuzzificação estão posicionados na entrada e saída do sistema de controle, respectivamente. Estes elementos são responsáveis por transformar as medidas obtidas dos sensores em conjuntos nebulosos (fuzzificador), e em transformar os conjuntos nebulosos obtidos na saída do controlador em valores não nebulosos de controle para o processo (defuzzificador)..

Observando a estrutura da Figura 11.1.1 aparecem os elementos fuzzificadores ArtInt, ArtNac e TraCom. O elemento defuzzificador Final e a propagação responsável pela aplicação das regras difusas. Todo o processo foi realizado utilizando o *software FuzzyTech 5.5*.



**Figura 11.1.1** - Estrutura do sistema difuso

**Fonte:** Estrutura gerada a partir software FuzzyTech

A figura a Figura 11.1.1 exhibe a estrutura de todo o sistema composto pelas interfaces de entrada, blocos de regras e interfaces de saída. As linhas de conexão simbolizam o fluxo de dados.

A estrutura do sistema identifica o fluxo de inferência difuso das variáveis de entrada para as variáveis de saída. A fuzzificação nas interfaces de entrada traduz entradas de dados linguísticos em valores difusos. A inferência difusa ocorre em blocos regra que contenham as regras de controle de linguística. A saída destes blocos de regra são variáveis linguísticas. A defuzzificação nas interfaces de saída converte-os em variáveis linguísticas. Segue o quadro do projeto estatístico difuso.

Variáveis de Entrada	3
Variáveis de Saída	1
Bloco de Regras	1
Regras	27
Funções de pertinência	14




**Quadro 11.1.1** - Projeto estatístico difuso

A definição das variáveis linguísticas acontece de forma subjetiva, assim como as regras difusas (funções de pertinência). Para representar os dados da produção científica foram utilizados os termos que indicassem a definição de produtividade. Variáveis linguísticas são usados para traduzir os valores reais em valores linguísticos. Os valores possíveis de uma variável linguística não são números, mas os chamados "termos linguísticos".

Para a classificação da produção dos estados os termos utilizados foram pequena, média e grande. Os mínimos e máximos utilizados na análise foram os mínimos e máximos encontrados na produção de artigos em periódicos nacionais, internacionais e em trabalhos completos em anais. Na variável classificação foi utilizada a escala de zero a dez e os pontos intermediários foram utilizados padrões do software por apresentar espaços equidistantes.

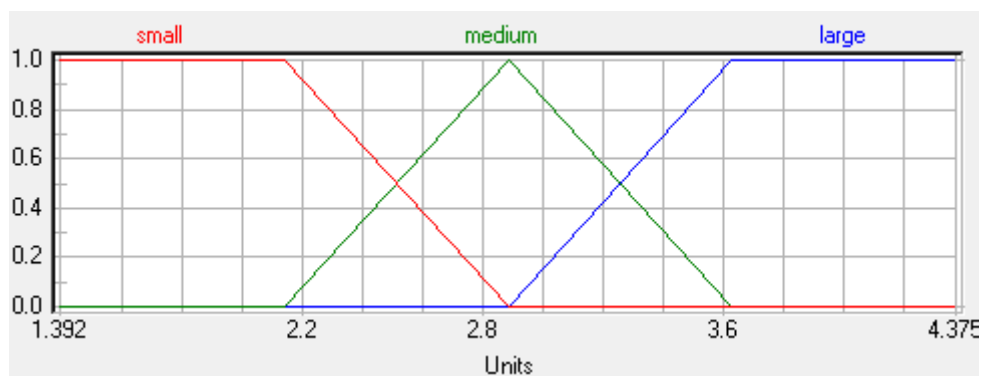
Por exemplo, para traduzir os valores da produção de artigos internacionais em variável linguística foram definidos três termos: pequena, média e grande. Dependendo do nível de produção de artigo internacional atual cada um desses termos foram escolhidos por descreverem essa produção mais ou menos bem. Cada termo é definido por uma função de pertinência ou método de fuzificação. Cada função define membros para qualquer valor da variável de entrada associado ao grau da adesão ao termo lingüístico.

Os quadros a seguir listam todas as variáveis do sistema, bem como o método de fuzificação e defuzificação respectivos. Também as propriedades de todas as variáveis base e os nomes estão listados.

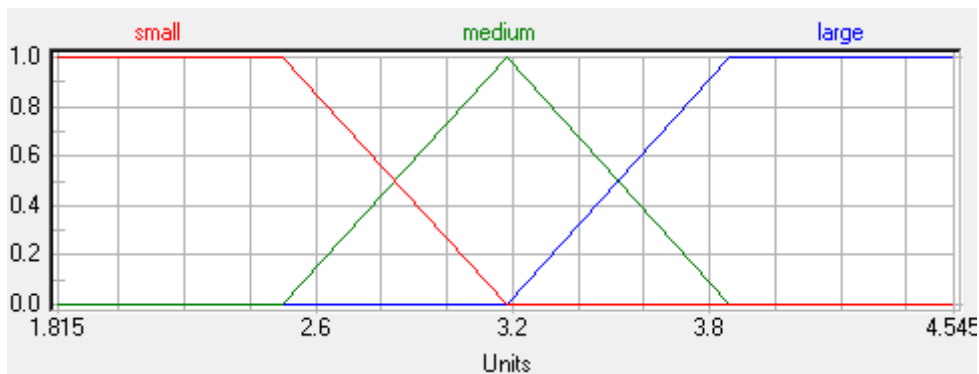
#	Variáveis	Métodos de Fuzzificação	Unit	Min	Max	Padrão	Termos
1	ARTINT	 - Compute MBF	Units	1.392	4.375	2.8835	pequena média grande
2	ARTNAC	 - Compute MBF	Units	1.815	4.545	3.18	pequena média grande
3	TRACOM	 - Compute MBF	Units	1.741	7.955	4.848	pequena média grande

**Quadro 11.1.2** – Variáveis de entrada

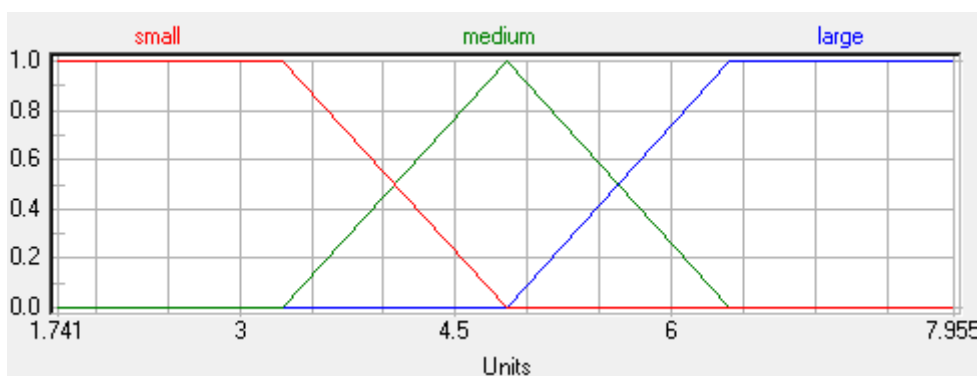
As funções de pertinência dos termos de uma variável linguística são normalmente exibidos em um gráfico. As figuras a seguir representam as funções de pertinência dos três termos de "produção". Foram utilizados valores padrão do software, ou seja, o software que definiu os pontos intermediários e convenientes por apresentarem-se equidistantes como os utilizados na definição da distância entre as faixas no estudo dos mapas temáticos.



**Figura 11.1.2** - Variáveis difusas para Artigos Internacionais - ARTINT

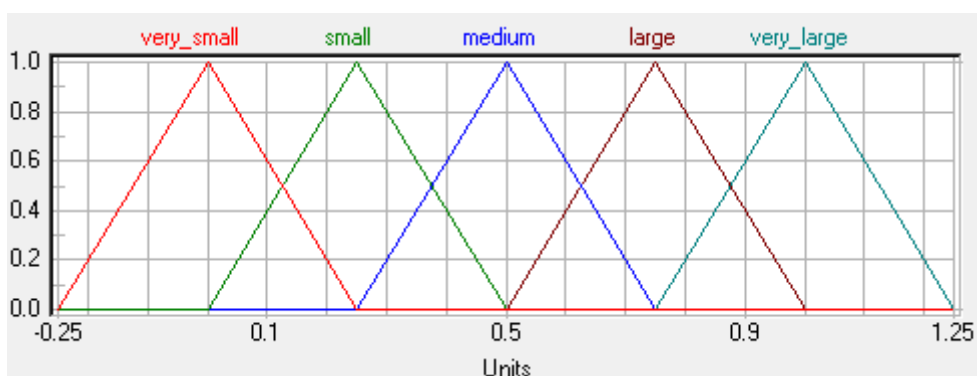


**Figura 11.1.3** - Variáveis difusas para Artigos Nacionais - ARTNAC




**Figura 11.1.4** - Variáveis difusas para Trabalho em Anais - TRACOM

Na definição das regras difusas, para cada variável foram utilizados os tipos de espaço trapezoidal para as extremidades. Para a saída, ou seja a defuzzificação, foi utilizado (muito\_baixa, baixa, média, alta e muito\_alta). Segue a representação desses dados.



**Figura 11.1.5** - Variáveis difusas para Classificação dos estados - FINAL

Diferentes métodos podem ser usados para a defuzzificação, mas neste estudo foi utilizado o método MoM (média de máxima) que condiciona o „resultado mais plausível“. O Quadro 11.1.3 contém a estrutura de defuzzificação com os cinco resultados possíveis de classificação da produção científica.

Variável	Método de Defuzzificação	Unit	Min	Max	Padrão	Termos
FINAL	 - média de máxima (MoM)	Units	-0.25	1.25	0.5	muito_pequena pequena média grande muito_grande

**Quadro 11.1.3** – Variável de saída

A base de regras difusas consiste em um conjunto de regras “IF-THEN”, o Quadro 11.1.4 apresenta a base de regras. As variáveis de entrada ARTINT, ARTNAC e TRACOM podem receber os seguintes valores: pequena, média, grande. A regra 1, por exemplo, estabelece que se todas as variáveis de entrada com o valor = pequena então a variável de saída FINAL assume o valor “muito\_baixa”. A cada regra pode ser estabelecido um “peso”, denominado Função de Certeza (DoS), normalmente 1.



SE			ENTÃO	
ARTINT	ARTNAC	TRACOM	DoS	FINAL
baixa	baixa	baixa	1.00	muito_baixa
baixa	baixa	média	1.00	muito_baixa
baixa	baixa	alta	1.00	média
média	baixa	baixa	1.00	muito_baixa
média	baixa	média	1.00	baixa
média	baixa	alta	1.00	média
alta	baixa	baixa	1.00	média
alta	baixa	média	1.00	média
alta	baixa	alta	1.00	muito_alta
baixa	média	baixa	1.00	muito_baixa
baixa	média	média	1.00	baixa
baixa	média	alta	1.00	média
média	média	baixa	1.00	baixa
média	média	média	1.00	média
média	média	alta	1.00	alta
alta	média	baixa	1.00	média
alta	média	média	1.00	alta
alta	média	alta	1.00	muito_alta
baixa	alta	baixa	1.00	média
baixa	alta	média	1.00	média
baixa	alta	alta	1.00	muito_alta
média	alta	baixa	1.00	média
média	alta	média	1.00	alta
média	alta	alta	1.00	muito_alta
alta	alta	baixa	1.00	muito_alta
alta	alta	média	1.00	muito_alta
alta	alta	alta	1.00	muito_alta

**Quadro 11.1.4** - Base de regras difusas

Para entender o sistema de classificação basta observar que para cada afirmação que um estado tem “alta” representa o grau de pertinência desse estado ao conjunto dos estados com produção alta. Isso permite que conjuntos como "alta" e "pequena" sejam tratados separadamente e válidas simultaneamente diferentemente da lógica clássica. O resultado pode ser visto no quadro abaixo .

UF	ARTINT	ARTNAC	TRACOM	FINAL	CLASSIFICAÇÃO
Acre	1,392	2,56	1,872	0	Muito Baixa
Alagoas	2,026	2,37	4,968	0	Muito Baixa
Amapá	2,852	1,815	1,741	0	Muito Baixa
Amazonas	2,948	2,503	2,936	0	Muito Baixa
Bahia	2,524	3,268	3,689	0,25	Baixa
Ceará	3,54	3,963	5,127	1	Muito_Alta
Distrito Federal	3,087	3,863	4,732	0,75	Alta
Espírito Santo	2,167	3,216	5,641	0,5	Média
Goiás	2,722	3,92	4,47	0,75	Alta
Maranhão	3,637	3,663	3,21	1	Muito_Alta
Mato Grosso	1,651	3,424	3,454	0	Muito Baixa
Mato Grosso do Sul	2,229	4,19	4,853	0,5	Média
Minas Gerais	3,306	4,545	5,077	1	Muito_Alta
Pará	2,414	2,859	4,255	0	Muito Baixa
Paraíba	2,421	4,233	7,955	1	Muito_Alta
Paraná	2,898	4,279	5,587	0,75	Alta
Pernambuco	2,619	3,723	5,395	0,75	Alta
Piauí	2,483	3,697	3,185	0,5	Média
Rio de Janeiro	3,716	3,222	4,575	0,75	Alta
Rio Grande do Norte	2,462	3,474	6,04	0,5	Média
Rio Grande do Sul	3,498	4,513	5,81	1	Muito_Alta
Rondônia	1,802	2,75	3,448	0	Muito Baixa
Roraima	2,032	4,455	2,136	0,5	Média
Santa Catarina	2,606	3,619	6,685	1	Muito_Alta
São Paulo	4,375	3,836	4,596	1	Muito_Alta
Sergipe	2,586	3,468	6,033	0,5	Média
Tocantins	2,286	4,267	3,008	0,5	Média

**Quadro 11.1.5 - Resultado difuso**

Os valores obtidos para a variável de saída FINAL são associados aos termos propostos: 0 = muito\_baixa, 0,25 = baixa, 0,5 = média, 0,75 = alta e 1 = muito\_alta. Com base nesses dados foi gerado o gráfico abaixo indicando a classificação dos estados segundo sua produção científica a partir de três indicadores.



Como mencionado anteriormente, dentro da lógica convencional, os termos podem ser apenas "verdadeiro" ou "falso". A lógica difusa permite uma generalização da lógica convencional, prevendo termo entre "verdadeiro" e "falso" como "quase verdade" ou "parcialmente falso". Portanto, o uso da lógica difusa permite a emulação dos dados por código especial e obter um resultado que represente a linguagem natural.

A justificativa para utilização da lógica difusa baseia-se em dois aspectos. Primeiro, lógica difusa é uma necessidade quando a interpretação da informação disponível é demasiada imprecisa para justificar apenas o uso de números, e segundo, quando há uma tolerância para a imprecisão que podem ser explorados para alcançar tratabilidade, robustez, solução de baixo custo, e melhor relacionamento com realidade (ZADEH, 1996).

Ainda seguindo o esforço de demonstrar as facilidades do uso de tecnologias para desenvolver os processos matemáticos e estatísticos neste estudo, esta a disposição no apêndice 2 o relatório de análise dos dados gerados pelo software FuzzyTech.

## 12 CONSIDERAÇÕES FINAIS

O Brasil sempre se mostrou um país de contrastes e diversidade marcantes. Talvez por suas dimensões continentais associadas a um terrível histórico de má distribuição de riquezas e recursos. Um país capaz de abrigar um “gigante” como o estado de São Paulo, que inegavelmente é responsável por uma expressiva fatia da pesquisa e desenvolvimento científico nacional, mas também abriga “heróis” que, nas mais remotas regiões do país, praticamente sozinhos e com poucos recursos, são capazes de superar esse “gigante” quando modificados os métodos de comparação. Analisando os mapas temáticos e histogramas a partir de índices *per capita* fica claro que a produção científica apresenta um quadro de similaridade muito grande com vários estados praticamente na mesma faixa de valores.

Numa rápida observação vai parecer ser óbvio que o número de doutores é uma grandeza diretamente proporcional ao número de publicações, mas isso não parece tão claro sem um estudo que proponha correlacionar tais variáveis. Este estudo apresentou uma demonstração que não existem tantas diferenças, com base em um índice de produtividade, na quantificação da produção científica nos estados. É importante pensar na melhor forma de disponibilizar recursos financeiros e humanos além de proporcionar e aprofundar discussões e de tentar erradicar velhos preconceitos.

É necessário também estudos mais profundos que levem em consideração o fator de impacto das publicações. E, além disso, é necessária a utilização de variáveis qualitativas apoiando dados quantitativos. Tais estudos devem caracterizar melhor cada estado e a importância desses no cenário nacional. Assim, será possível perceber, de forma mais significativa, onde os recursos aplicados em ciência e tecnologia têm obtido melhores resultados.

No papel de fator de produção e desenvolvimento, a informação, leva o bibliotecário moderno a uma mudança de postura profissional. Novas técnicas e tecnologias assim como o surgimento de novos canais de comunicação têm tornado o trabalho do bibliotecário mais complexo, levando-o a necessidade de adquirir novos conhecimentos e desenvolver novas habilidades. Essas mudanças no campo

da Biblioteconomia levam-na a se distanciar cada vez mais de ser uma atividade tão somente técnica.

Este estudo deve contribuir tanto para a Ciência da Informação quanto para a Biblioteconomia por demonstrar a importância de se utilizar métodos e procedimentos adequados à natureza do estudo que se quer desenvolver. E também, o quanto é possível e produtivo se ampliar as possibilidades de análise de dados sobre os produtos e serviços desenvolvidos nas bibliotecas e outros serviços de informação. Por fim, deve alargar os horizontes da Ciência da Informação para pesquisas semelhantes.

### 13 REFERÊNCIAS

ALCAÍN, M. D.; SAN MILLÁN, M. J. Uso y tendencias de las técnicas bibliométricas bibliométricas em en ciencias sociales y humanas a nivel internacional. **Revista Española de Documentación Científica**, v. 16, n. 1, 1993.

ARAÚJO, Vânia Maria Rodrigues Hermes de; FREIRE, Isa Maria. Conhecimento para o desenvolvimento: reflexões para o profissional da informação. **Informação & Sociedade: Estudos** - v.9, n.1, 1999.

BARBOSA, R. R. Perspectivas profissionais e educacionais em biblioteconomia e ciência da informação. **Ci. Inf.**, Brasília, vol.27, n.1, 1998.

Bolaño, César Ricardo Siqueira; Melo, Ricardo Oliveira Lacerda de. Tecnologias da informação e da comunicação e desenvolvimento regional. **Revista de Economia Política das Tecnologias da Informação e Comunicação**, v. 2, n.2, p. 63-82, jul/ago 2000.

BORKO, H. Information science: what is it? *American Documentation*, Jan. 1968.

BRASIL, Ministério da Ciência e Tecnologia. Indicadores: apresentação/histórico. Disponível em: <<http://www.mct.gov.br/index.php/content/view/740.html>>. Acesso em: 20 jun. 2009.

BRISOLLA, Sandra Negraes. Indicadores para apoio à tomada de decisão. *Revista . Ci. Inf.*, Brasília, v. 27, n. 2, p. 221-225, maio/ago. 1998.

BUFREM, Leilah; PRATES, Yara. O saber científico registrado e as práticas de mensuração da informação. **Ci. Inf.**, Brasília, v. 34, n. 2, p. 9-25, maio/ago. 2005.

BURNHAM, Teresinha Fróes. Sociedade da informação, sociedade do conhecimento, sociedade da aprendizagem. In: LUBISCO, Nídia M. L.; RANDÃO, Lídia M. B. **Informação e informática**. Salvador: EDUFBA, 2000. p. 283-307.

Burnham, Terezinha Fróes. Da sociedade da informação à sociedade da aprendizagem: cidadania e participação sócio-política (in) formação do trabalhador. In **Proceedings CIFORM - Encontro Nacional de Ciência da Informação VI**, Salvador - Bahia. 2005

CARVALHO, Érica Lílian de. **A utilização e interpretação de mapas temáticos no ensino de geografia como forma de compreender a realidade**. Disponível em: <<http://artigos.netsaber.com.br/>>. Acesso em: 10 set 2009.

CASTELLS, Manuel. **A Era da Informação**: economia, sociedade e cultura, vol. 1, São Paulo: Paz e terra, 1999.

CASTRO, C. M. Há produção científica no Brasil. **Ciência e Cultura**, v. 37, n. 7, p. 165-187, 1985. Suplemento.

CHEN, Sherry Y.; LIU, Xiaohui. The contribution of data mining to information science. **Journal of Information Science**, v. 30, n. 6, p. 550–558 2004.

CHERNOFF, H. The use of faces to represent points in k-dimensional space. **Journal of the American Statistical Association**, 68(342):361-368, 1973.

CHOO, Chun Wei. A organização do conhecimento: uma visão holística de como as organizações usam a informação. In: \_\_\_\_\_. **A organização do conhecimento: como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões**. São Paulo: Senac, 2000. Cap. 1, p.27-61.

FAPESP. **Indicadores de ciência, tecnologia e inovação em São Paulo 2004**. São Paulo: FAPESP, 2005.

FARIAS, Alfredo Alves de. **Estatística Aplicada**. 2 ed. São Paulo: Saraiva, 2002.

FONSECA, Edson Nery. **Bibliometria: teoria e prática**. São Paulo: Cultrix, 1986.

FREITAS, Carla Maria Dal Sasso et al. Introdução à Visualização de Informações. **Revista de Informática Teórica e Aplicada**, Porto Alegre - RS, v. 8, n. 2, p. 143-158, 2001.

FREITAS, Carla Maria Dal Sasso et al. Introdução à Visualização de Informações. **Revista de Informática Teórica e Aplicada**, Porto Alegre - RS, v. 8, n. 2, p. 143-158, 2001.

GOHN, Maria da Glória. **Movimentos sociais e luta pela moradia**. São Paulo: Loyola, 1991. 191p.

GONSALVES, Carlos Roberto Colares; BARROS, Helena Maria Motta de Oliveira. Indicadores de c&t no Brasil: situação atual e perspectivas. **Anais... IV TALLER IBEROAMERICANO/INTERAMERICANO de indicadores de c&t conacyt - México**, 12-14 julho 1999.

GUIMARAES, REINALDO. Pesquisa no Brasil: a reforma tardia. **São Paulo Perspec.**, São Paulo, v. 16, n. 4, out. 2002.

KONDO, E. K. Desenvolvendo indicadores estratégicos em ciência e tecnologia: as principais questões. **Ci. Inf.**, Brasília, v. 27, n. 2, p. 128-133, maio/ago. 1998.

LE COADIC. **A ciência da informação**. Brasília, D.F.: Brique de Lemos, 1996.

LEE, Michael D. et al. Visualizations of binary data: A comparative evaluation. **Int. J. Human-Computer Studies**, v. 59, p. 569–602, 2003.

LEE, Yun-Seok; KOO, Han-Suh; JEONG, Chang-Sung. A straight line detection using principal component analysis. **Pattern Recognition Letters**, v. 27, p. 1744–1754, 2006.

MACÍAS-CHAPULA, C. A. O papel da informetria e da cientometria e sua perspectiva nacional e internacional. **Ci. Inf.**, Brasília, v. 27, n. 2, p. 134-140, maio/ago. 1998.

MARTINS, Gilberto De Andrade. **Estatística Geral e Aplicada**. 3 ed. São Paulo: Atlas, 2005.



MARTINS, Gilberto De Andrade; DONAIRE, Denis. **Princípios de Estatística**. 1 ed. São Paulo: Atlas, 1981.

MEADOWS, A. J. **A comunicação científica**. Brasília: Briquet de Lemos, 2000.

MOITA NETO, J. M. ; SANTOS JÚNIOR, José Ribeiro dos . Uso de Estatística Multivariada no Ensino de Química. **Spin Ensino e Pesquisa**, Belo Horizonte, v. 3, n. 5, p. 33-35, 1996.

MOITA NETO, J. M. Estatística multivariada: uma visão didática-metodológica. **Crítica Revista de Filosofia e Ensino**, v. 1, n. 1, p. 1-1, 2004. Disponível em: <[http://criticanarede.com/cien\\_estatistica.html](http://criticanarede.com/cien_estatistica.html)>. Acesso em 10.03.2010.

MOITA NETO, José Machado; MOITA, Graziella Ciaramella. Uma introdução à análise exploratória de dados multivariados. **Quím. Nova**, São Paulo, v. 21, n. 4, 1998. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-40421998000400016&lng=pt&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-40421998000400016&lng=pt&nrm=iso)>. Acesso em: 02 Out. 2006.

MOITA NETO, José Machado; MOITA, Graziella Ciaramella. Uma introdução à análise exploratória de dados multivariados. **Quím. Nova**, São Paulo, v. 21, n. 4, 1998. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-40421998000400016&lng=pt&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-40421998000400016&lng=pt&nrm=iso)>. Acesso em: 02 Out. 2006.

PORCARO, R M. Indicadores da sociedade atual - informação, conhecimento, inovação e aprendizado intensivos. A perspectiva da OECD. **DataGramZero**, v. 6 n. 4. ago. 2005.

PRAT, Anna Maria. Avaliação da produção científica como instrumento para o desenvolvimento da ciência e da tecnologia

ROCHA, Elisa Maria Pinto; FERREIRA, Marta Araújo Tavares. Indicadores de ciência, tecnologia e inovação: mensuração dos sistemas de CTel nos estados brasileiros. **Ci. Inf.**, Brasília, v. 33, n. 3, p. 61-68, set./dez. 2004.

ROUSSEAU, R. Indicadores bibliométricos e econométricos para a avaliação de instituições científicas. **Ci. Inf.**, Brasília, v. 27, n. 2, p. 149-158, maio/ago. 1998.

RUAS, Maria das Graças. Análise de políticas públicas: conceitos básicos. In.: O estudo da política: tópicos selecionados. Brasília: Paralelo 15, 1998.

SAIÃO, L. F. Bases de dados: a metáfora da memória científica. **Ci. Inf.**, Brasília, v. 25, n.3, 1996.

SANCHEZ, Oscar Adolfo. O poder burocrático e o controle da informação. *Lua Nova*, n. 58, 2003. P. 89-120.

SANTOS, Carla. **Estatística Descritiva**: Manual de auto-aprendizagem. Lisboa: Silabo, 2007.

SANTOS, Raimundo Nonato Macedo dos; KOBASHI, Nair Yumiko. bibliometria, cientometria, infometria: conceitos e aplicações. **Pesquisa brasileira em Ciência da Informação**, Brasília, v.2, n.1, p.155-172, jan./dez. 2009.

SILVA, CMSC da; ARRUDA, GM. A formação do profissional de biblioteconomia frente às novas tendências do mercado globalizado. In: ENCONTROS BIBLI, 6, UFSC - Florianópolis, 1998. 1 arquivo (68 Kb).Word for Windows 98.

SPINAK, Ernesto. Indicadores cientometricos. **Ci. Inf.**, Brasília, v. 27, n. 2, 1998.

TARAPANOV, K. **Técnicas para tomada de decisão nos sistemas de informação**. Brasília: Thesaurus, 1995.

TRZESNIAK, Piotr. Indicadores quantitativos: reflexões que antecedem seu estabelecimento. **Ci. Inf.**, Brasília, v. 27, n. 2, 1998.

UNIVERSIDADE ESTADUAL DE MARINGÁ. Lógica Difusa. Disponível em: <[http://www.din.uem.br/ia/control/fuz\\_prin.htm](http://www.din.uem.br/ia/control/fuz_prin.htm)>. Acesso em: 10 set 2006.

VALENTIM, ML (org.). Formação do profissional da informação. São Paulo: Polis; 2002.

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

ZADEH, Lofti A. Stochastic Finite-State Systems in Control Theory. **Proceedings Symposium on Optimal Control and Nonlinear Systems**, Faculte des Sciences de Paris, 123-143, 1963.

ZADEH, Lofti A. Fuzzy Sets. **Information and control**, v. 8, 338-353, 1965.

ZADEH, Lofti A. The concept of a linguistic variable and its application to approximate reasoning. **Information Sciences**, v. 8, n. 9 p. 199–249,301–357, 1975.

ZADEH, Lofti A. A theory of approximate reasoning. In J. Hayes, D. Michie, and L. Mikulich, editors, **Machine Intelligence**, v.9, p. 149–194. Halstead, New York, 1979.

ZADEH, Lofti A. A computational approach to fuzzy quantifiers in natural languages. **Comput. and Math. with Applications**, v. 9, p. 149–184, 1983.

ZADEH, Lofti A. Syllogistic reasoning in fuzzy logic and its application to usuality and reasoning with dispositions. **IEEE Transactions on Systems, Man, and Cybernetics**, v. 15, n. 6, p. 754–763, Nov./Dec. 1985.

ZADEH, Lofti A. Fuzzy Logic = Computing with Words. **IEEE Transactions On Fuzzy Systems**, v. 4, n. 2, maio 1996.

WORMELL, I. Informetrics: exploring databases as analytical tools. **Database**, v. 21, n. 5, p. 25-30, Out./Nov. 1998.

# APÊNDICES

## Apêndice 1 – Relatório gerado com os dados da análise dos componentes principais do software XLSTAT

XLSTAT 2009.3.02 - Análise de Componentes Principais (ACP) - em 30/05/2011 às 16:20:11

Tabela de observações/variáveis: Documento = Tabela principal de trabalho.xls / Planilha = Transformação de variáveis / Intervalo = 'Transformação de variáveis'!\$C\$10:\$J\$37 / 27 linhas e 8 colunas

Nomes das observações: Documento = Tabela principal de trabalho.xls / Planilha = Transformação de variáveis / Intervalo 'Transformação de variáveis'!\$B\$10:\$B\$37 / 27 linhas e 1 coluna

Tipo de ACP: Pearson (n)

Tipo de biplot: Biplot de correlação / Coeficiente = Automático

Estatísticas descritivas:

Variável	Observações	Obs. com dados faltantes	Obs. sem dados faltantes	Mínimo	Máximo	Média	Desvio padrão
Artigos nacional	27	0	27	-2,433	1,408	0,000	1,000
Artigos internacional	27	0	27	-1,876	2,478	0,000	1,000
Trabalhos completos ar	27	0	27	-1,812	2,326	0,000	1,000
Livros	27	0	27	-2,351	2,407	0,000	1,000
Capítulos de livros	27	0	27	-2,796	1,833	0,000	1,000
Outras publicações	27	0	27	-1,833	2,104	0,000	1,000
Resumos Periódicos	27	0	27	-1,465	2,423	0,000	1,000
Resumo Anais	27	0	27	-2,324	2,047	0,000	1,000

Matriz de correlação (Pearson (n)):

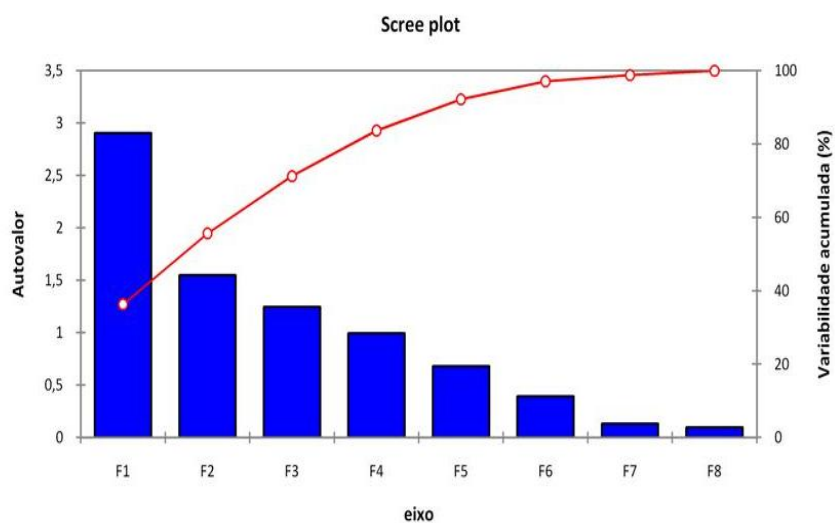
Variáveis	ARTNAC	ARTINT	TRACOM	Livros	CAPLIV	OUTPUB	RESPER	RESANA
Artigos nacional	<b>1</b>	0,289	<b>0,415</b>	0,072	0,354	<b>0,493</b>	0,356	0,334
Artigos internacional	0,289	<b>1</b>	0,214	0,357	0,332	-0,177	<b>0,761</b>	<b>0,622</b>
Trabalhos completos ar	<b>0,415</b>	0,214	<b>1</b>	0,210	<b>0,462</b>	-0,299	0,174	0,214
Livros	0,072	0,357	0,210	<b>1</b>	0,152	-0,028	-0,022	0,267
Capítulos de livros	0,354	0,332	<b>0,462</b>	0,152	<b>1</b>	0,141	0,157	0,067
Outras publicações	<b>0,493</b>	-0,177	-0,299	-0,028	0,141	<b>1</b>	-0,136	-0,097
Resumos Periódicos	0,356	<b>0,761</b>	0,174	-0,022	0,157	-0,136	<b>1</b>	<b>0,517</b>
Resumo Anais	0,334	<b>0,622</b>	0,214	0,267	0,067	-0,097	<b>0,517</b>	<b>1</b>

Os valores em negrito são diferentes de 0 com um nível de significância  $\alpha=0,05$

Análise de Componentes Principais:

Autovalores:

	F1	F2	F3	F4	F5	F6	F7	F8
Autovalor	2,905	1,548	1,248	0,994	0,682	0,394	0,133	0,098
Variabilidade (%)	36,312	19,347	15,595	12,420	8,520	4,920	1,665	1,221
% acumulada	36,312	55,659	71,253	83,674	92,194	97,114	98,779	100,000



Autovetores:

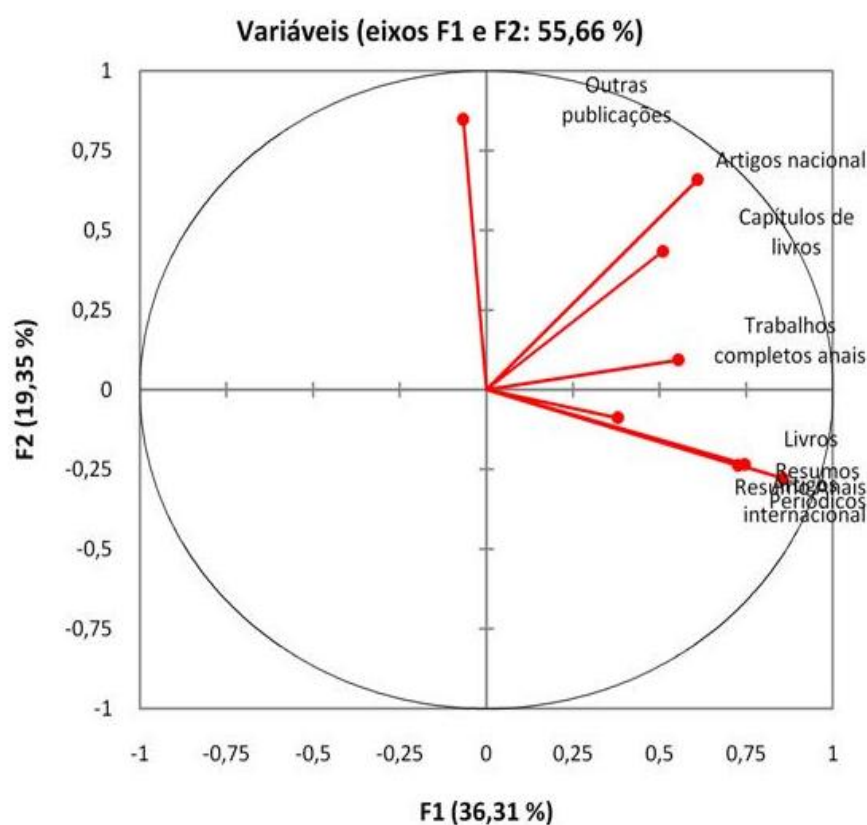
	F1	F2	F3	F4	F5	F6	F7	F8
Artigos nacional	0,358	0,530	-0,126	-0,095	-0,383	-0,237	-0,453	-0,398
Artigos internacional	0,502	-0,223	-0,166	0,088	0,344	-0,120	-0,516	0,513
Trabalhos completos ar	0,325	0,075	0,612	-0,261	-0,412	-0,116	0,244	0,451
Livros	0,223	-0,070	0,288	0,842	0,008	-0,289	0,164	-0,207
Capítulos de livros	0,299	0,349	0,402	-0,136	0,626	0,383	0,058	-0,258
Outras publicações	-0,039	0,682	-0,373	0,251	0,076	0,025	0,334	0,462
Resumos Periódicos	0,438	-0,188	-0,366	-0,308	0,164	-0,399	0,556	-0,221
Resumo Anais	0,426	-0,191	-0,256	0,171	-0,374	0,725	0,135	-0,057

Cargas fatoriais:

	F1	F2	F3	F4	F5	F6	F7	F8
Artigos nacional	0,609	0,660	-0,141	-0,095	-0,316	-0,149	-0,165	-0,124
Artigos internacional	0,856	-0,278	-0,186	0,088	0,284	-0,076	-0,188	0,160
Trabalhos completos ar	0,555	0,094	0,683	-0,261	-0,341	-0,073	0,089	0,141
Livros	0,379	-0,088	0,321	0,839	0,006	-0,182	0,060	-0,065
Capítulos de livros	0,509	0,434	0,449	-0,136	0,517	0,240	0,021	-0,081
Outras publicações	-0,067	0,848	-0,416	0,250	0,063	0,016	0,122	0,144
Resumos Periódicos	0,746	-0,234	-0,408	-0,307	0,136	-0,250	0,203	-0,069
Resumo Anais	0,726	-0,237	-0,286	0,171	-0,309	0,455	0,049	-0,018

Correlações entre as variáveis e fatores:

	F1	F2	F3	F4	F5	F6	F7	F8
Artigos nacional	0,609	0,660	-0,141	-0,095	-0,316	-0,149	-0,165	-0,124
Artigos internacional	0,856	-0,278	-0,186	0,088	0,284	-0,076	-0,188	0,160
Trabalhos completos ar	0,555	0,094	0,683	-0,261	-0,341	-0,073	0,089	0,141
Livros	0,379	-0,088	0,321	0,839	0,006	-0,182	0,060	-0,065
Capítulos de livros	0,509	0,434	0,449	-0,136	0,517	0,240	0,021	-0,081
Outras publicações	-0,067	0,848	-0,416	0,250	0,063	0,016	0,122	0,144
Resumos Periódicos	0,746	-0,234	-0,408	-0,307	0,136	-0,250	0,203	-0,069
Resumo Anais	0,726	-0,237	-0,286	0,171	-0,309	0,455	0,049	-0,018



Contribuições das variáveis (%):

	F1	F2	F3	F4	F5	F6	F7	F8
Artigos nacional	12,788	28,103	1,582	0,903	14,655	5,631	20,481	15,856
Artigos internacional	25,244	4,981	2,766	0,775	11,807	1,449	26,619	26,360
Trabalhos completos ar	10,593	0,565	37,398	6,837	17,013	1,353	5,932	20,309
Livros	4,955	0,496	8,276	70,886	0,006	8,377	2,703	4,302
Capítulos de livros	8,932	12,180	16,152	1,857	39,236	14,645	0,341	6,657
Outras publicações	0,155	46,491	13,893	6,306	0,584	0,062	11,183	21,326
Resumos Periódicos	19,175	3,542	13,364	9,499	2,704	15,918	30,929	4,868
Resumo Anais	18,158	3,643	6,569	2,936	13,996	52,564	1,812	0,322

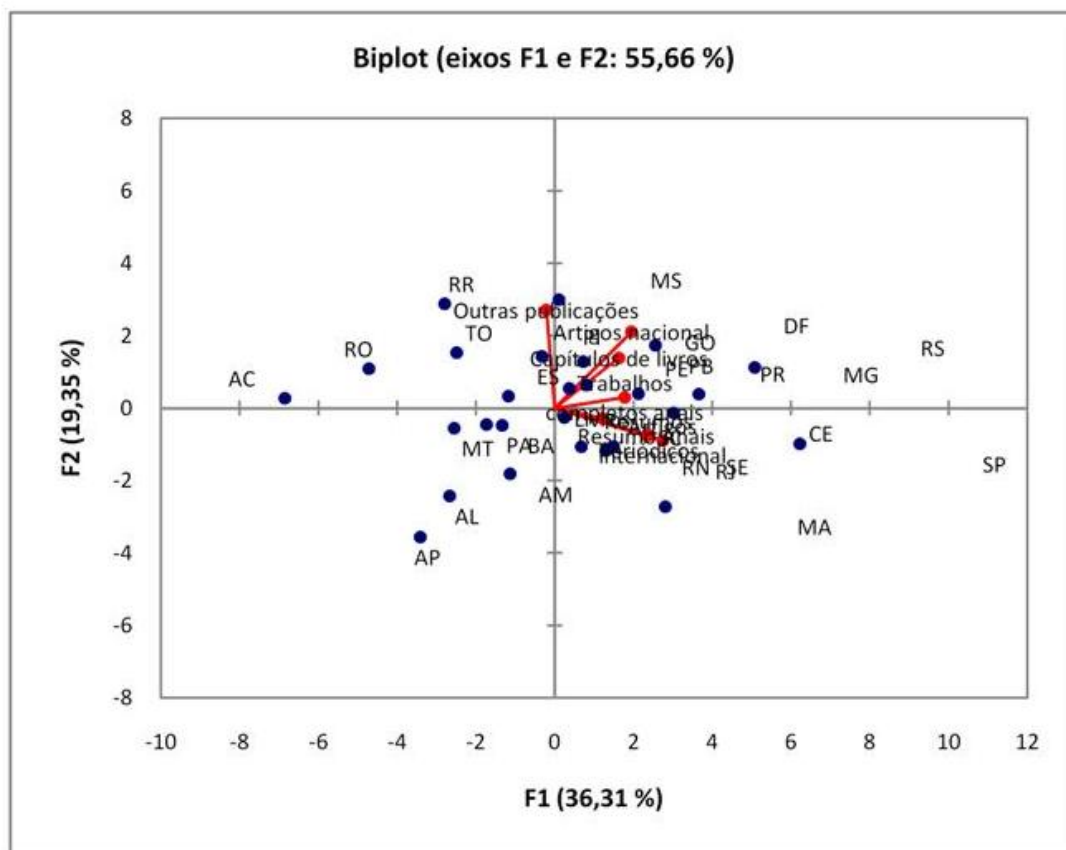
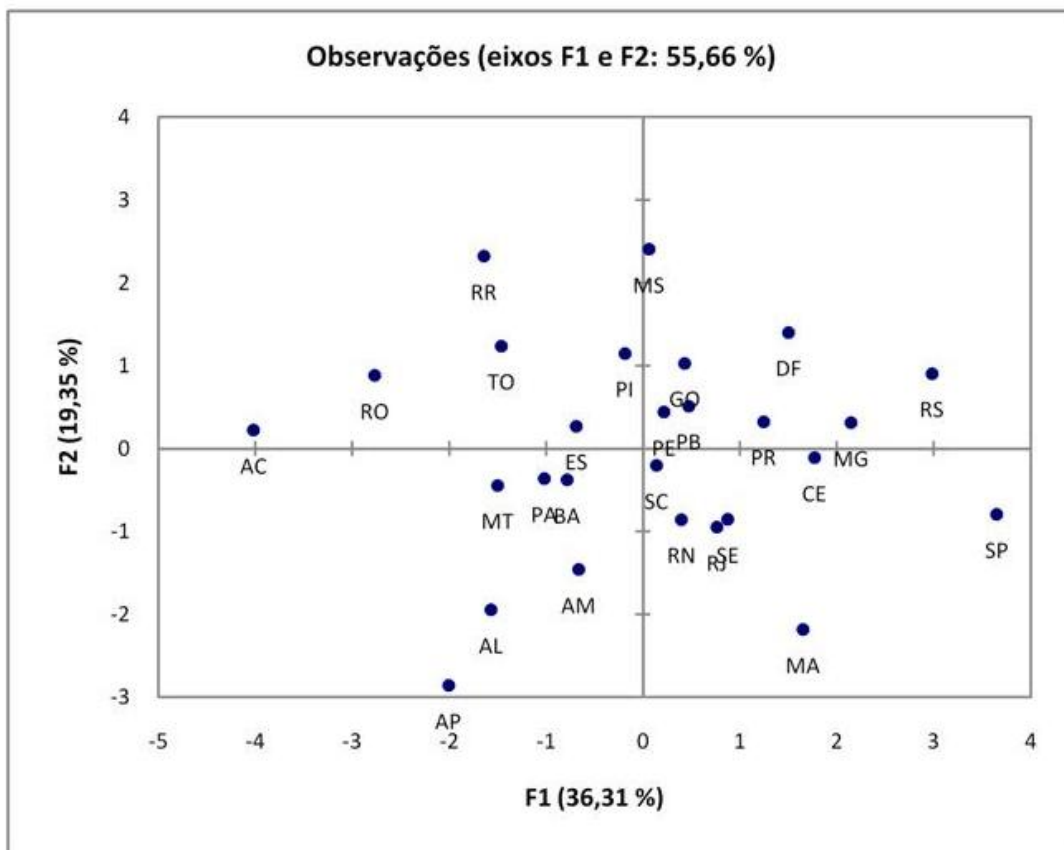
Cossenos quadrados das variáveis:

	F1	F2	F3	F4	F5	F6	F7	F8
Artigos nacional	0,371	0,435	0,020	0,009	0,100	0,022	0,027	0,015
Artigos internacional	0,733	0,077	0,035	0,008	0,080	0,006	0,035	0,026
Trabalhos completos ar	0,308	0,009	0,467	0,068	0,116	0,005	0,008	0,020
Livros	0,144	0,008	0,103	0,704	0,000	0,033	0,004	0,004
Capítulos de livros	0,259	0,189	0,202	0,018	0,267	0,058	0,000	0,007
Outras publicações	0,004	0,720	0,173	0,063	0,004	0,000	0,015	0,021
Resumos Periódicos	0,557	0,055	0,167	0,094	0,018	0,063	0,041	0,005
Resumo Anais	0,527	0,056	0,082	0,029	0,095	0,207	0,002	0,000

Coordenadas dos fatores:

Observação	F1	F2	F3	F4	F5	F6	F7	F8
AC	-4,018	0,222	-0,797	-1,177	0,413	0,830	-0,137	0,029
AL	-1,568	-1,946	1,022	-1,104	-0,124	-0,057	0,434	-0,171
AP	-2,006	-2,859	-1,302	2,409	-0,190	-0,340	0,188	0,628
AM	-0,665	-1,460	-0,275	0,602	0,859	0,498	-0,057	0,094
BA	-0,783	-0,378	-0,027	-0,539	0,525	-0,192	-0,145	-0,301
CE	1,769	-0,109	-0,372	-0,582	-0,062	0,862	-0,563	0,378
DF	1,502	1,400	1,055	2,062	1,262	-0,217	0,344	-0,234
ES	-0,688	0,269	1,345	-0,947	0,101	0,741	0,077	0,095
GO	0,429	1,029	-0,314	-0,114	0,339	-0,157	0,226	0,039
MA	1,651	-2,183	-2,176	-1,014	-0,076	-0,431	-0,106	-0,473
MT	-1,500	-0,444	0,217	0,723	-0,841	-0,005	0,118	-1,070
MS	0,061	2,406	-0,045	0,853	-0,260	0,274	0,497	0,264
MG	2,146	0,311	-0,547	0,021	-0,462	-0,206	-0,217	-0,280
PA	-1,016	-0,364	1,032	0,031	0,658	0,580	-0,160	-0,061
PB	0,474	0,512	1,532	-1,767	-1,287	-0,807	-0,095	0,393
PR	1,246	0,324	-0,387	-0,302	-0,906	-0,409	0,100	0,100
PE	0,215	0,441	0,283	-1,081	0,139	-0,272	0,186	0,122
PI	-0,187	1,148	-1,872	-0,469	0,457	-0,209	0,915	0,158
RJ	0,763	-0,949	0,957	0,247	1,429	-0,702	-0,809	0,186
RN	0,397	-0,857	0,298	-0,904	-1,152	0,246	0,399	0,097
RS	2,982	0,903	0,549	0,421	0,328	0,419	-0,132	-0,295
RO	-2,768	0,881	1,118	-0,156	1,439	-0,243	0,032	-0,044
RR	-1,639	2,319	-1,924	0,268	-0,749	0,799	-0,765	-0,005
SC	0,141	-0,203	2,009	0,429	-0,560	-0,940	-0,163	0,134
SP	3,651	-0,795	-1,240	-0,490	1,166	0,343	0,173	0,157
SE	0,876	-0,855	1,156	1,780	-1,741	1,228	0,021	0,052
TO	-1,462	1,237	-1,296	0,801	-0,707	-1,631	-0,358	0,008





Contribuições das observações (%):

	F1	F2	F3	F4	F5	F6	F7	F8
AC	20,583	0,118	1,884	5,166	0,929	6,476	0,525	0,031
AL	3,134	9,063	3,102	4,541	0,083	0,031	5,231	1,109
AP	5,130	19,559	5,029	21,638	0,197	1,089	0,979	14,969
AM	0,564	5,099	0,225	1,351	4,011	2,333	0,090	0,334
BA	0,783	0,343	0,002	1,082	1,499	0,345	0,588	3,438
CE	3,991	0,029	0,410	1,262	0,021	6,988	8,806	5,433
DF	2,875	4,693	3,306	15,843	8,650	0,445	3,283	2,085
ES	0,604	0,173	5,370	3,345	0,055	5,168	0,165	0,344
GO	0,234	2,534	0,293	0,049	0,625	0,232	1,414	0,058
MA	3,475	11,408	14,059	3,831	0,031	1,751	0,314	8,472
MT	2,869	0,472	0,140	1,949	3,845	0,000	0,386	43,433
MS	0,005	13,849	0,006	2,713	0,367	0,705	6,864	2,645
MG	5,874	0,232	0,887	0,002	1,158	0,399	1,309	2,981
PA	1,317	0,317	3,165	0,004	2,355	3,170	0,714	0,142
PB	0,286	0,628	6,963	11,641	8,993	6,132	0,249	5,861
PR	1,980	0,251	0,444	0,339	4,461	1,578	0,279	0,381
PE	0,059	0,466	0,237	4,355	0,105	0,696	0,963	0,560
PI	0,045	3,153	10,403	0,820	1,136	0,409	23,286	0,948
RJ	0,741	2,154	2,720	0,227	11,098	4,643	18,202	1,313
RN	0,201	1,758	0,263	3,047	7,214	0,568	4,418	0,354
RS	11,336	1,952	0,894	0,659	0,584	1,655	0,487	3,308
RO	9,771	1,859	3,708	0,090	11,256	0,554	0,028	0,072
RR	3,427	12,872	10,988	0,269	3,047	6,002	16,272	0,001
SC	0,025	0,099	11,985	0,685	1,706	8,322	0,741	0,682
SP	16,991	1,511	4,565	0,897	7,391	1,108	0,836	0,941
SE	0,978	1,749	3,964	11,805	16,471	14,181	0,012	0,102
TO	2,724	3,659	4,985	2,390	2,712	25,019	3,562	0,002

Cossenos quadrados das observações:

	F1	F2	F3	F4	F5	F6	F7	F8
AC	0,846	0,003	0,033	0,073	0,009	0,036	0,001	0,000
AL	0,281	0,433	0,119	0,139	0,002	0,000	0,022	0,003
AP	0,198	0,403	0,084	0,286	0,002	0,006	0,002	0,019
AM	0,110	0,531	0,019	0,090	0,184	0,062	0,001	0,002
BA	0,417	0,097	0,001	0,197	0,187	0,025	0,014	0,062
CE	0,649	0,002	0,029	0,070	0,001	0,154	0,066	0,030
DF	0,198	0,172	0,098	0,373	0,140	0,004	0,010	0,005
ES	0,124	0,019	0,473	0,235	0,003	0,144	0,002	0,002
GO	0,119	0,685	0,064	0,008	0,074	0,016	0,033	0,001
MA	0,199	0,348	0,346	0,075	0,000	0,014	0,001	0,016
MT	0,461	0,040	0,010	0,107	0,145	0,000	0,003	0,234
MS	0,001	0,829	0,000	0,104	0,010	0,011	0,035	0,010
MG	0,856	0,018	0,056	0,000	0,040	0,008	0,009	0,015
PA	0,341	0,044	0,352	0,000	0,143	0,111	0,008	0,001
PB	0,027	0,031	0,278	0,371	0,196	0,077	0,001	0,018
PR	0,534	0,036	0,051	0,031	0,282	0,058	0,003	0,003
PE	0,028	0,119	0,049	0,716	0,012	0,045	0,021	0,009
PI	0,006	0,213	0,566	0,036	0,034	0,007	0,135	0,004
RJ	0,102	0,158	0,161	0,011	0,359	0,087	0,115	0,006
RN	0,047	0,219	0,026	0,244	0,396	0,018	0,047	0,003
RS	0,841	0,077	0,028	0,017	0,010	0,017	0,002	0,008
RO	0,647	0,066	0,105	0,002	0,175	0,005	0,000	0,000
RR	0,197	0,395	0,272	0,005	0,041	0,047	0,043	0,000
SC	0,004	0,007	0,731	0,033	0,057	0,160	0,005	0,003
SP	0,772	0,037	0,089	0,014	0,079	0,007	0,002	0,001
SE	0,073	0,069	0,127	0,300	0,288	0,143	0,000	0,000
TO	0,230	0,165	0,181	0,069	0,054	0,287	0,014	0,000

## Apêndice 2 – Relatório gerado com os dados de análise da lógica Fuzzy do software FUZZYTECH

### 1 General Information

Author: LEVI SANTOS  
 Created: quarta-feira, 6 de julho de 2011  
 Print Date: quarta-feira, 6 de julho de 2011

#### Edition

Edition Name: fuzzyTECH 5.54d Professional Edition  
 Neuro Modul: NeuroFuzzy add-on Module installed

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Compute MBF	Compute Membership Function (Fuzzification Method)
MoM	Mean of Maximum (Defuzzification Methode)
BSUM	Bounded Sum Fuzzy Operator for Result Aggregation
MIN	Fuzzy Operator for AND Aggregation
MAX	Fuzzy Operator for OR Aggregation
GAMMA	Compensatory Operator for Aggregation
PROD	Fuzzy Operator for Composition
LV	Linguistic Variable
MBF	Membership Function
RB	Rule Block

**FUZZY\_FINAL**

## Project Description

Input Variables	3
Output Variables	1
Intermediate Variables	0
Rule Blocks	1
Rules	27
Membership Functions	14

Table 1: Project Statistics

## System Structure

The system structure identifies the fuzzy logic inference flow from the input variables to the output variables. The fuzzification in the input interfaces translates analog inputs into fuzzy values. The fuzzy inference takes place in rule blocks which contain the linguistic control rules. The output of these rule blocks are linguistic variables. The defuzzification in the output interfaces translates them into analog variables.

The following figure shows the whole structure of this fuzzy system including input interfaces, rule blocks and output interfaces. The connecting lines symbolize the data flow.

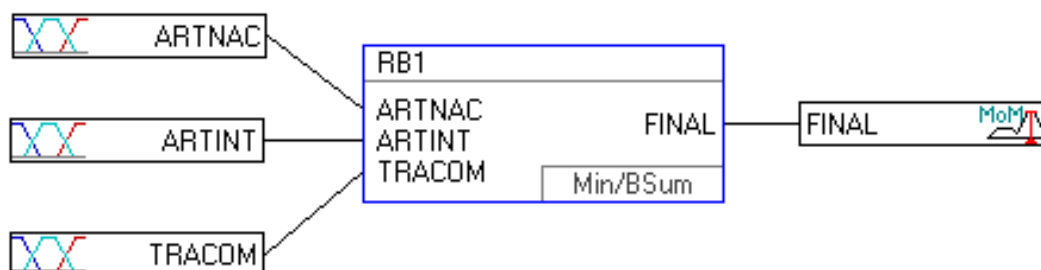


Figure 1: Structure of the Fuzzy Logic System

## Variables

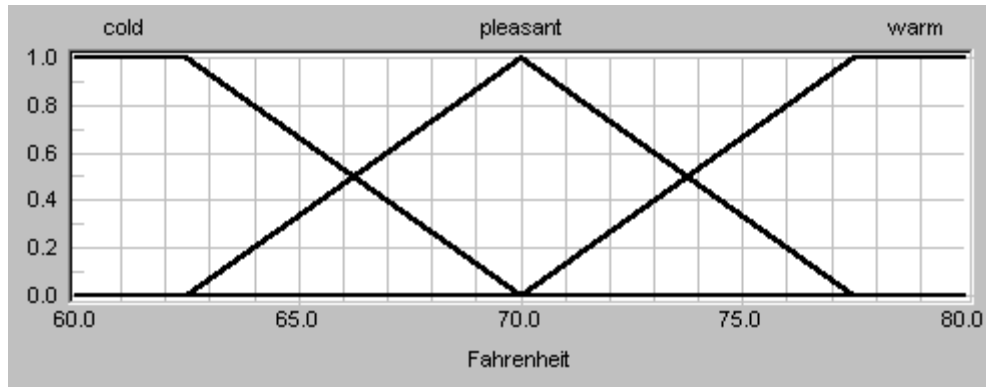
This chapter contains the definition of all linguistic variables and of all membership functions.

Linguistic variables are used to translate real values into linguistic values. The possible values of a linguistic variable are not numbers but so called 'linguistic terms'.

For example:

To translate the real variable 'temperature' into a linguistic variable three terms, 'cold', 'pleasant' and 'warm' are defined. Depending on the current temperature level each of these terms describes the 'temperature' more or less well. Each term is defined by a membership function (MBF). Each membership function defines for any value of

the input variable the associated degree of membership of the linguistic term. The membership functions of all terms of one linguistic variable are normally displayed in one graph. The following figure plots the membership functions of the three terms for the example 'temperature'.



*Membership Function of 'temperature'*

A 'temperature' of 64 °F is a member of the MBFs for the terms:

cold           to the degree of 0.8  
pleasant       to the degree of 0.2  
warm           to the degree of 0.0

Linguistic variables have to be defined for all input, output and intermediate variables. The membership functions are defined using a few definition points only.

The following tables list all variables of the system as well as the respective fuzzification or defuzzification method. Also the properties of all base variables and the term names are listed.


## Inputs

#	Variable Name	Type	Unit	Min	Max	Media	Term Names
1	ARTINT		Units	1.392	4.375	2.8835	small medium large
2	ARTNAC		Units	1.815	4.545	3.18	small medium large
3	TRACOM		Units	1.741	7.955	4.848	small medium large

*Table 2: Variables of Group "Inputs"*

Fuzzification Methods  
 Compute MBF  
 Categorical Variable

Look up MBF  
 Display

 Fuzzy Input

### Outputs








#	Variable Name	Type	Unit	Min	Max	Mediat	Term Names
4	FINAL		Units	-0.25	1.25	0.5	very_small small medium large very_large

Table 3: Variables of Group "Outputs"

Defuzzification Methods

-  Center of Maximum (CoM)
-  Center of Area (CoA)
-  Fuzzy Output

-  Mean of Maximum (MoM)
-  Hyper CoM
-  Force

The default value of an output variable is used if no rule is firing for this variable. Different methods can be used for the defuzzification, resulting either into the 'most plausible result' or the 'best compromise'.

The 'best compromise' is produced by the methods:

- CoM (Center of Maximum)
- CoA (Center of Area)
- CoA BSUM, a version especially for efficient VLSI implementations

The 'most plausible result is produced by the methods:

- MoM (Mean of Maximum)
- MoM BSUM, a version especially for efficient VLSI implementations

### Input Variable "ARTINT"

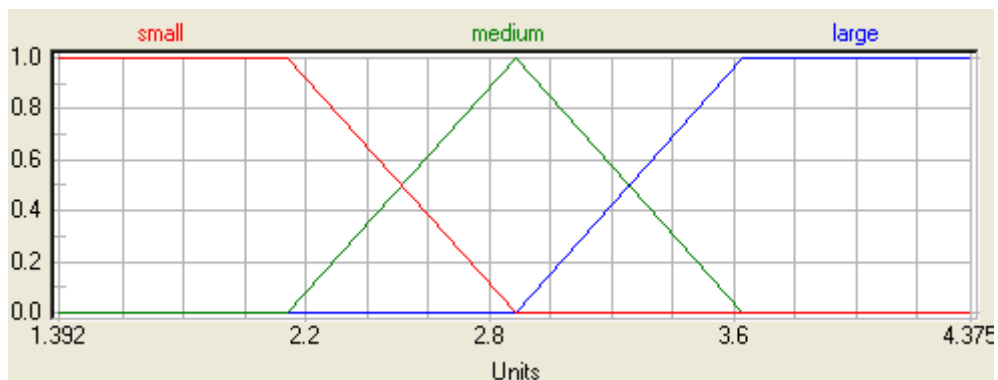


Figure 2: MBF of "ARTINT"

Term Name	Shape/Par.	Definition Points (x, y)		
small	linear	(1.392, 1)	(2.13775, 1)	(2.8835, 0)
		(4.375, 0)		
medium	linear	(1.392, 0)	(2.13775, 0)	(2.8835, 1)
		(3.62925, 0)	(4.375, 0)	
large	linear	(1.392, 0)	(2.8835, 0)	(3.62925, 1)
		(4.375, 1)		

Table 4: Definition Points of MBF "ARTINT"

### Input Variable "ARTNAC"

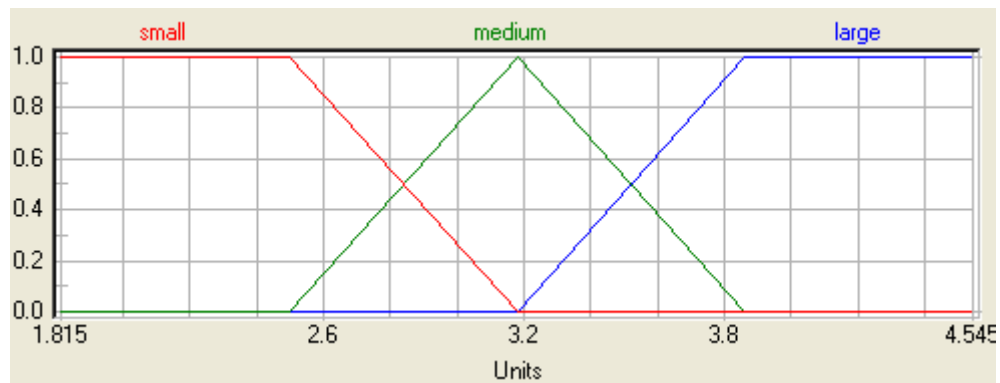


Figure 3: MBF of "ARTNAC"

Term Name	Shape/Par.	Definition Points (x, y)		
small	linear	(1.815, 1)	(2.4975, 1)	(3.18, 0)
		(4.545, 0)		
medium	linear	(1.815, 0)	(2.4975, 0)	(3.18, 1)
		(3.8625, 0)	(4.545, 0)	
large	linear	(1.815, 0)	(3.18, 0)	(3.8625, 1)
		(4.545, 1)		

Table 5: Definition Points of MBF "ARTNAC"

### Input Variable "TRACOM"



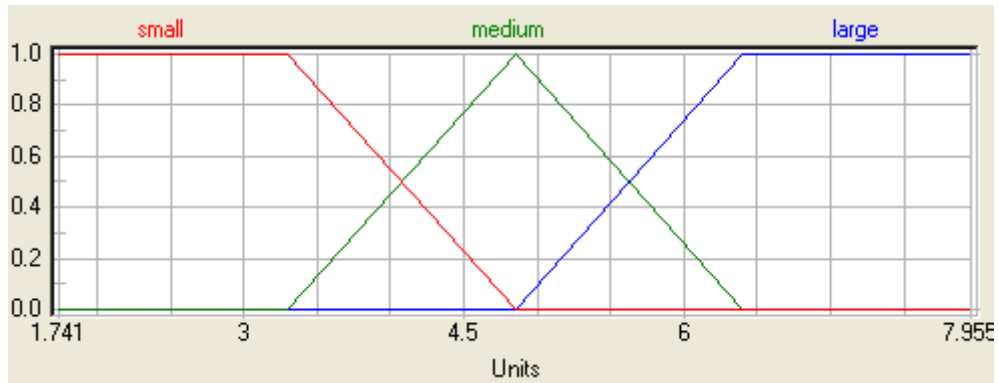


Figure 4: MBF of "TRACOM"

Term Name	Shape/Par.	Definition Points (x, y)
small	linear	(1.741, 1) (3.2945, 1) (4.848, 0) (7.955, 0)
medium	linear	(1.741, 0) (3.2945, 0) (4.848, 1) (6.4015, 0) (7.955, 0)
large	linear	(1.741, 0) (4.848, 0) (6.4015, 1) (7.955, 1)

Table 6: Definition Points of MBF "TRACOM"

**Output Variable "FINAL"**

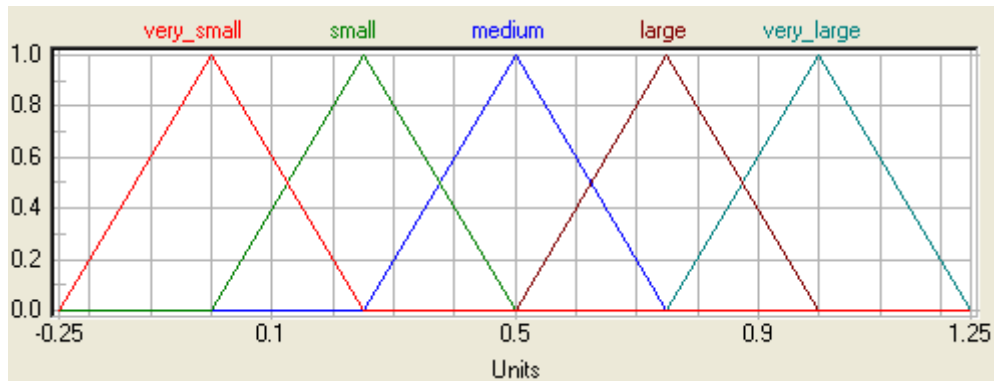


Figure 5: MBF of "FINAL"

Term Name	Shape/Par.	Definition Points (x, y)
very_small	linear	(-0.25, 0) (0, 1) (0.25, 0) (1.25, 0)
small	linear	(-0.25, 0) (0, 0) (0.25, 1) (0.5, 0) (1.25, 0)
medium	linear	(-0.25, 0) (0.25, 0) (0.5, 1) (0.75, 0) (1.25, 0)
large	linear	(-0.25, 0) (0.5, 0) (0.75, 1) (1, 0) (1.25, 0)
very_large	linear	(-0.25, 0) (0.75, 0) (1, 1) (1.25, 0)

Term Name	Shape/Par.	Definition Points (x, y)
		(1.25, 0)

Table 7: Definition Points of MBF "FINAL"

Rule Blocks

The rule blocks contain the control strategy of a fuzzy logic system. Each rule block confines all rules for the same context. A context is defined by the same input and output variables of the rules.

The rules' 'if' part describes the situation, for which the rules are designed. The 'then' part describes the response of the fuzzy system in this situation. The degree of support (DoS) is used to weigh each rule according to its importance.

The processing of the rules starts with calculating the 'if' part. The operator type of the rule block determines which method is used. The operator types MIN-MAX, MIN-AVG and GAMMA are available. The characteristic of each operator type is influenced by an additional parameter.

For example:

- MIN-MAX, parameter value 0 = Minimum Operator (MIN)
- MIN-MAX, parameter value 1 = Maximum Operator (MAX)
- GAMMA, parameter value 0 = Product Operator (PROD)

The minimum operator is a generalization of the Boolean 'and'; the maximum operator is a generalization of the Boolean 'or'.

The fuzzy composition eventually combines the different rules to one conclusion. If the BSUM method is used all firing rules are evaluated, if the MAX method is used only the dominant rules are evaluated.

**Rule Block "RB1"**

**Parameter**

- Aggregation: MINMAX
- Parameter: 0.00
- Result Aggregation: BSUM
- Number of Inputs: 3
- Number of Outputs: 1
- Number of Rules: 27

IF			THEN	
ARTINT	ARTNAC	TRACOM	DoS	FINAL
small	small	small	1.00	very_small
small	small	medium	1.00	very_small
small	small	large	1.00	medium

IF			THEN	
medium	small	small	1.00	very_small
medium	small	medium	1.00	small
medium	small	large	1.00	medium
large	small	small	1.00	medium
large	small	medium	1.00	medium
large	small	large	1.00	very_large
small	medium	small	1.00	very_small
small	medium	medium	1.00	small
small	medium	large	1.00	medium
medium	medium	small	1.00	small
medium	medium	medium	1.00	medium
medium	medium	large	1.00	large
large	medium	small	1.00	medium
large	medium	medium	1.00	large
large	medium	large	1.00	very_large
small	large	small	1.00	medium
small	large	medium	1.00	medium
small	large	large	1.00	very_large
medium	large	small	1.00	medium
medium	large	medium	1.00	large
medium	large	large	1.00	very_large
large	large	small	1.00	very_large
large	large	medium	1.00	very_large
large	large	large	1.00	very_large

Table 8: Rules of the Rule Block "RB1"

## Settings

### Global Options

Base Variable Data Type: Double Precision  
 Computation Options: Fast CoA

### Code Generator Options

Settings: Public Input and Output  
 Comments  
 Compatibility: ANSI C

### Online Options

Communication Channel: OCC (Online Communication Channel): D:\Arquivos de programas\INFORM\fuzzyTECH 5.5\Ftocc.dll  
 Refresh Time: 55 ms  
 Timeout: 1100 ms

# **ANEXOS**

## Anexo 1 – Portal do Plano Tabular\*

The screenshot shows the 'Plano Tabular' web portal. At the top, there is a blue header with the text 'Censos' and 'Plano Tabular - Diretório de Grupos de Pesquisa no Brasil'. To the right of the header are logos for 'CNPq' (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and 'GrupoStela'. Below the header is a yellow navigation bar with links: 'Grupos', 'Linhas de pesquisa', 'Pesquisadores', 'Estudantes', 'Técnicos', 'Produção C, T&A', 'Grupos/Empresas', and 'Empresas/Grupos'. The main content area is titled 'Grupos de Pesquisa'. It features a dropdown menu labeled 'Escolha a variável de filtro:'. Below this is a list of variables with 'Adicionar' and 'Excluir' buttons. A double-headed arrow indicates the list is scrollable. Below the list is another dropdown menu labeled 'Escolha o ano do censo:' with '2008' selected. At the bottom is a 'Consultar' button. To the right of the interface, there is a text block explaining the module's purpose and providing instructions on how to use the filters and search function.

**Grupos de Pesquisa**

Escolha a variável de filtro:

Adicionar Excluir

Escolha o ano do censo: 2008

Consultar

Este módulo permite visualizar quantitativamente o perfil da pesquisa no Brasil. Trata-se de um sistema que oferece a possibilidade de cruzamento de variáveis capazes de gerar um número muito grande de tabelas, que podem ser salvas em planilhas ou em arquivos htm ou html para futuras consultas.

**Instruções**

Selecione uma das unidades de análise disponíveis, localizadas na barra horizontal superior: GRUPOS, LINHAS DE PESQUISA, PESQUISADORES, ESTUDANTES, TÉCNICOS, PRODUÇÃO C,T&A, GRUPOS/EMPRESAS ou EMPRESAS/GRUPOS. Em seguida, na janela vertical à esquerda, escolha uma ou mais variáveis de filtro e selecione, uma a uma, as opções da variável clicando no botão "Adicionar". Ao final, pressione o botão "Consultar". As setas azuis podem ser utilizadas para alterar a ordem das opções selecionadas. Para excluir uma variável selecionada, clique sobre ela e em seguida no botão Excluir. Dependendo da consulta, a tela seguinte poderá ainda apresentar variáveis de conteúdo, que o usuário deverá selecionar e, em seguida, pressionar novamente o botão "Consultar". Num dos filtros disponíveis, ano do censo, o usuário poderá selecionar qual o Censo que deseja consultar (o padrão é sempre o último censo publicado).

\* <http://www.plano...>



# Plano Tabular

## Diretório de Grupos de Pesquisa no Brasil

- saiba mais
- exemplos
- pág inicial



Grupos | Linhas de pesquisa | Pesquisadores | Estudantes | Técnicos | Produção C, T&A | Grupos/Empresas | Empresas/Grupos

### Produção CT&A

Escolha a variável de filtro:

Geográfico/Institucional ▾

- Brasil
- Região
- UF
- Instituição

Adicionar

Excluir

Por UF

Escolha o ano do censo:

2008 ▾

Consultar

Escolha a variável de conteúdo:

### Tipo de produção

- Produção bibliográfica
- Produção técnica
- Orientação concluída
- Produção artística/cultural
- Todos

### Autores

- Pesquisadores
- Estudantes
- Pesquisadores doutores
- Todos

Consultar

Produção bibliográfica segundo uf para pesquisadores doutores, 2005-2008, Censo 2008.\*

UF	Total de autores	Artigos completos publicados em periódicos especializados		Trabalhos completos publicados em anais de eventos	Livros ou capítulos de livro publicados		Outras publicações bibliográficas (3)	Resumos de trabalhos publicados em(4)	
		Circulação nacional (1)	Circulação internacional (2)		Livros	Capítulos de livros		Periódicos especializados	Anais de eventos
Acre	125	320	174	234	23	179	698	3	808
Alagoas	567	1.344	1.149	2.817	160	835	1.820	87	4.137
Amapá	54	98	154	94	22	33	256	6	458
Amazonas	966	2.418	2.848	2.836	326	1.578	4.410	136	8.124
Bahia	3.040	9.935	7.672	11.215	896	5.225	14.234	501	22.142
Ceará	1.504	5.960	5.324	7.711	434	2.910	7.621	290	14.814
Distrito Federal	2.009	7.761	6.202	9.507	942	4.885	12.926	305	16.098
Espírito Santo	753	2.422	1.632	4.248	209	1.545	3.584	65	5.810
Goiás	1.367	5.359	3.721	6.111	437	2.619	8.457	266	10.865
Maranhão	410	1.502	1.491	1.316	115	527	1.582	157	3.992
Mato Grosso	766	2.623	1.265	2.646	276	1.056	3.456	73	6.154
Mato Grosso do Sul	1.088	4.559	2.425	5.280	385	2.124	8.206	114	8.978
Minas Gerais	6.992	31.778	23.117	35.499	2.433	12.552	37.260	1.770	66.079
Pará	1.023	2.925	2.470	4.353	329	1.995	4.682	74	7.737
Paraíba	1.565	6.624	3.789	12.449	430	2.650	7.078	260	11.308
Paraná	5.324	22.782	15.431	29.746	1.728	8.341	29.212	1.224	47.301
Pernambuco	2.588	9.636	6.779	13.963	730	4.888	13.633	520	19.524
Piauí	406	1.501	1.008	1.293	111	675	2.902	108	3.257
Rio de Janeiro	9.200	29.640	34.190	42.092	3.400	18.503	37.513	1.565	64.942
Rio Grande do Norte	1.035	3.596	2.548	6.251	302	1.544	4.469	205	9.450
Rio Grande do Sul	6.491	29.292	22.705	37.713	2.502	15.622	35.713	1.443	63.013
Rondônia	116	319	209	400	35	233	650	4	595
Roraima	154	666	313	329	39	222	1.160	5	1.287
Santa Catarina	2.933	10.615	7.642	19.807	1.159	4.978	12.657	308	20.784
São Paulo	21.014	80.803	91.929	96.572	6.830	45.787	107.170	8.090	218.998
Sergipe	515	1.786	1.332	3.107	219	792	2.318	26	5.577
Tocantins	266	1.135	608	800	91	278	1.785	35	1.709
<b>TOTAIS</b>	<b>72.271</b>	<b>277.219</b>	<b>248.127</b>	<b>358.189</b>	<b>24.563</b>	<b>142.556</b>	<b>365.452</b>	<b>17.640</b>	<b>643.921</b>

**ANEXO 2** – Versão anterior de acesso aos indicadores por unidade da Federação no portal do MCT

MCT - Indicadores Estaduais de C&T

Página 1 de 1



Formulário de levantamento de  
investimentos municipais em c&t

(clique no botão direito do mouse e salve )

INDICADORES NACIONAIS de  
CIÊNCIA & TECNOLOGIA



Ministério da Ciência e Tecnologia - MCT

**Indicadores Estaduais de Ciência e Tecnologia**

---

**Indicadores Consolidados**

- Recursos aplicados
- Recursos humanos
- Bolsas
- Produção científica
- Patentes

---

Assessoria de Acompanhamento e Avaliação - ASCAV / Coordenação-Geral de Indicadores

**:: Produção científica ::**

Tabela 4.1 ::	Produção científica no diretório dos grupos de pesquisa do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), por unidade da federação e regiões <table border="1" data-bbox="501 546 858 600"> <tr> <td data-bbox="501 546 683 573">1998-2001</td> <td data-bbox="692 546 858 573">2000-2003</td> </tr> <tr> <td data-bbox="501 573 683 600">parte 1 parte 2</td> <td data-bbox="692 573 858 600">parte 1 parte 2</td> </tr> </table>	1998-2001	2000-2003	parte 1 parte 2	parte 1 parte 2
1998-2001	2000-2003				
parte 1 parte 2	parte 1 parte 2				
Tabela 4.2 ::	Produção técnica no diretório dos grupos de pesquisa do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), por unidade da federação e regiões <table border="1" data-bbox="501 680 858 734"> <tr> <td data-bbox="501 680 683 707">1998-2001</td> <td data-bbox="692 680 858 707">2000-2003</td> </tr> <tr> <td data-bbox="501 707 683 734">parte 1 parte 2</td> <td data-bbox="692 707 858 734">parte 1 parte 2</td> </tr> </table>	1998-2001	2000-2003	parte 1 parte 2	parte 1 parte 2
1998-2001	2000-2003				
parte 1 parte 2	parte 1 parte 2				
Tabela 4.3 ::	Orientações concluídas segundo seus tipos por unidade da federação e regiões <table border="1" data-bbox="501 792 858 846"> <tr> <td data-bbox="501 792 683 819">1998-2001</td> <td data-bbox="692 792 858 819">2000-2003</td> </tr> <tr> <td data-bbox="501 819 683 846">parte 1 parte 2</td> <td data-bbox="692 819 858 846">parte 1 parte 2</td> </tr> </table>	1998-2001	2000-2003	parte 1 parte 2	parte 1 parte 2
1998-2001	2000-2003				
parte 1 parte 2	parte 1 parte 2				