AFINIDADE
FUNDAMENTO AXIOMÁTICO
E
APLICAÇÕES

HELOISA BONVINO

Orientador
Prof.Dr. Pushpa Narayan Rathie

Dissertação apresentada no Instituto de Matemática, Estatística e Ciência da Computação da Universidade Esta - dual de Campinas como requisito parcial para o obtenção do Título de Mestre em Estatística.

Campinas - 1979.

A meus pais que me ensinaram o amor aos livros, que sempre me deram incentivo, apoio e compreensão.

ÍNDICE

INTRODUÇÃO

A Teoria da Informação teve origem na termodinâmica estatís ca, mais precisamente na Engenharia de Comunicação.

Hartley (1928), um importante precursor, definiu uma medida de informação, que é um caso particular da entropia, definida por Shannon (1948). Simultaneamente e independente de Shannon , Wiener (1948) obteve o mesmo resultado para a entropia.

Por esta razão, os mesmos são considerados os precursores mais importantes no desenvolvimento da Teoria da Informação.

Fisher (1948) definiu uma medida de quantidade de informação, contida nos dados, sobre um parâmetro desconhecido, que é usada em Inferência Estatística, na estimação de parâmetros. Esta medida foi o ponto inicial da introdução da Teoria da Informação, na Estatística Matemática.

Muitos resultados obtidos no desenvolvimento da Teoria da Informação são dados por Ash (1965), Rényi (1970), Kullback (1959).

As aplicações que inicialmente se situavam no campo da Teoria da Comunicação, são, hoje, encontradas em muitas outras áreas, abrangendo desde a Estatística até a Psicologia, ver Kullback (1959), Attneave (1959), Theil (1967).

Dada a grande variedade de medidas, na Teoria da Informação, e do vasto campo de aplicações das mesmas, fizemos um estudo sobre uma destas medidas, denominada Afinidade. Nosso trabalho consta de uma parte teórica, onde definimos propriedades e teoremas, para duas ou mais distribuições discretas; e, uma parte aplicada, onde usamos dados reais de consumo de energia elétrica, nas capitais administrativas, do Estado de São Paulo, e ainda apresentamos outras aplicações possíveis.

O primeiro capítulo foi divido em duas secções. Na primeira, apresentamos definição e propriedades da medida de Afinidade entre k $(k > 2)$ distribuições discretas; e na segunda apresentamos definição e propriedades da Afinidade entre duas distribuições discretas, sendo que, neste caso omitimos as demonstrações, por serem análogas às encontradas na secção anterior.

No segundo capítulo, damos uma definição axiomática para a medida em questão, considerando, aqui também, o caso de k $(k > 2)$ distribuições discretas e o caso em que temos apenas duas.

Apresentamos os resultados para o caso em que temos apenas duas distribuições discretas, separadamente do geral, porque são estes resultados que usamos na parte prática.

No capítulo seguinte mostramos uma das possíveis aplicações práticas da Afinidade, usando, para tal, o consumo de energia elétrica pelas diversas classes, nas regiões administrativas do Estado de São Paulo. Medindo a Afinidade em nove anos consecutivos, podemos observar, através da tendência, se está havendo alguma mudança dentro das classe.

No último capítulo enumeramos uma série de outros resultados e aplicações que envolvem a medida de Afinidade.

CAPÍTULO I

DEFINIÇÕES E PROPRIEDADES

Neste primeiro capítulo, definiremos e estabeleceremos propriedades para a Medida de Afinidade, entre distribuições discretas, levando em consideração o número de populações envolvidas.

A Afinidade foi estudada por Matusita ((1954),(1955),(1957)), juntamente com uma outra medida, a distância. Mais informações sobre os resultados de afinidade e outras medidas relacionadas podem ser encontradas no livro de Mathai e Rathie (1975).

1.1 - AFINIDADE ENTRE VÁRIAS POPULAÇÕES

1.1.1 - DEFINIÇÃO

Sejam $P_j = (p_{j1},\ldots,p_{jn})$, $j = 1,\ldots,k$, $\sum_{i=1}^{n} p_{ji} = 1$, $p_{ji} \geq 0$, $j = 1,\ldots,k$ e $i = 1,\ldots,n$; k distribuições discretas. A Afinidade entre as distribuições $P_1,\ldots,P_k$, denotada por $\rho_{n,k}$, é definida por:

$$(1.1.1)\quad \rho_{n,k} \equiv \rho_{n,k}(p_{11},\ldots,p_{1n} : p_{21},\ldots,p_{2n}:\ldots:p_{k1},\ldots,p_{kn}) =$$

$$= \sum_{i=1}^{n} \left[\prod_{j=1}^{k} p_{ji} \right]^{1/k} \quad ; \quad k \geq 2$$

1.1.2 - PROPRIEDADES

Apresentaremos nesta secção, as principais propriedades da medida de Afinidade entre k distribuições ($k > 2$).

i) RECURSIVIDADE

Sejam $P_j = (p_{j1},\ldots,p_{jn})$, $j = 1,\ldots k$; $p_{ji} \geq 0$, $i=1,\ldots,n$; $j = 1,\ldots,k$ e $\sum_{i=1}^{n} p_{ji} = 1$, $j = 1,\ldots,k$.

(1.1.2) $$\rho_{n,k}(p_{11},\ldots,p_{1n} : p_{21},\ldots,p_{2n}:\ldots:p_{k1},\ldots,p_{kn}) =$$

$$= \rho_{n-1,k}(p_{11}+p_{12},p_{13},\ldots,p_{1n} : p_{21}+p_{22},p_{23},\ldots,p_{2n}:\ldots:p_{k1}+p_{k2},p_{k3},\ldots,p_{kn}) +$$

$$+ K(p_{11},p_{12} : p_{21},p_{22} : \ldots : p_{k1},p_{k2}) \cdot$$

$$\cdot \rho_{2,k}\left(\frac{p_{11}}{p_{11}+p_{12}}, \frac{p_{12}}{p_{11}+p_{12}} : \frac{p_{21}}{p_{21}+p_{22}}, \frac{p_{22}}{p_{21}+p_{22}}:\ldots:\frac{p_{k1}}{p_{k1}+p_{k2}}, \frac{p_{k2}}{p_{k1}+p_{k2}}\right) ,$$

$$\forall\ n > 2,\ \sum_{i=1}^{2} p_{ji} > 0 \ ,\ j=1,\ldots,k .$$

onde:

(1.1.3) $$K(p_{11},p_{12} : p_{21},p_{22}:\ldots:p_{k1},p_{k2}) =$$

$$= \frac{\left\{ \sum_{i=1}^{2} \left[\prod_{j=1}^{k} p_{ji} \right]^{1/k} - \left[\prod_{j=1}^{k} \left(\sum_{i=1}^{2} p_{ji} \right) \right]^{1/k} \right\} \left[\prod_{j=1}^{k} \left(\sum_{i=1}^{2} p_{ji} \right) \right]^{1/k}}{\sum_{i=1}^{2} \left[\prod_{j=1}^{k} p_{ji} \right]^{1/k}}$$

PROVA:

$$\rho_{n,k}(p_{11},\ldots,p_{1n}:p_{21},\ldots,p_{2n}:\ldots:p_{k1},\ldots,p_{kn}) -$$

$$- \rho_{n-1,k}(p_{11}+p_{12},p_{13},\ldots,p_{1n}:p_{21}+p_{22},p_{23},\ldots,p_{2k}:\ldots:p_{k1}+p_{k2},p_{k3},\ldots,p_{kn}) =$$

$$= \sum_{i=1}^{n} [\prod_{j=1}^{k} p_{ji}]^{1/k} - \{[\prod_{j=1}^{k} (\sum_{i=1}^{2} p_{ji})]^{1/k} + \sum_{i=3}^{n} (\prod_{j=1}^{k} p_{ji})^{1/k}\} =$$

$$= \sum_{i=1}^{2} [\prod_{j=1}^{k} p_{ji}]^{1/k} - [\prod_{j=1}^{k} (\sum_{i=1}^{2} p_{ji})]^{1/k} =$$

$$= \frac{\{\sum_{i=1}^{2} [\prod_{j=1}^{k} p_{ji}]^{1/k} - [\prod_{j=1}^{k} (\sum_{i=1}^{2} p_{ji})]^{1/k}\} [\prod_{j=1}^{k} (\sum_{i=1}^{2} p_{ji})]^{1/k}}{\sum_{i=1}^{2} [\prod_{j=1}^{k} p_{ji}]^{1/k}} \cdot$$

$$\cdot \frac{\sum_{i=1}^{2} [\prod_{j=1}^{k} p_{ji}]^{1/k}}{[\prod_{j=1}^{k} (\sum_{i=1}^{2} p_{ji})]^{1/k}}$$

$$= K(p_{11},p_{12}:p_{21},p_{22}:\dots:p_{k1},p_{k2}) \cdot$$

$$\cdot \rho_{2,k}\left(\frac{p_{11}}{p_{11}+p_{12}}, \frac{p_{12}}{p_{11}+p_{12}}: \frac{p_{21}}{p_{21}+p_{22}}, \frac{p_{22}}{p_{21}+p_{22}}:\dots:\frac{p_{k1}}{p_{k1}+p_{k2}}, \frac{p_{k2}}{p_{k1}+p_{k2}}\right)$$

$$\forall\ n > 2\ ,\ \sum_{i=1}^{2} p_{ji} > 0,\ j=1,\dots,k$$

Portanto, temos (1.1.2).

ii) SIMETRIA

A medida de Afinidade $\rho_{n,k}$ é simétrica em k-uplas $\{p_{1i}, p_{2i},\dots,p_{ki}\}$, $i = 1,2,\dots,n$. Ou seja, (1.1.4) $\rho_{n,k}(p_{11},\dots,p_{1n}:p_{21},\dots,p_{2n}:\dots:p_{k1},\dots,p_{kn}) =$

$= \rho_{n,k}(p_{1a_1},\dots,p_{1a_n}:p_{2a_1},\dots,p_{2a_n}:\dots:p_{ka_1},\dots,p_{ka_n})$,

onde $\{a_1,\dots,a_n\}$ é uma permutação arbitrária de $\{1,\dots,n\}$.

iii) SIMETRIA EM DISTRIBUIÇÃO

Seja

(1.1.5) $\rho_{n,k}(p_{11},\dots,p_{1n}:p_{21},\dots,p_{2n}:\dots:p_{k1},\dots,p_{kn}) =$

$= \rho_{n,k}(P_1 : P_2:\dots: P_k)$.

$\rho_{n,k}$ é simétrica em distribuição, ou seja, $\rho_{n,k}(P_1: P_2:\dots:P_k) = \rho_{n,k}(P_{a_1}: P_{a_2}:\dots: P_{a_k})$, onde $\{a_1, a_2,\dots,a_k\}$ é uma permutação arbitrária de $\{1,2,\dots,k\}$.

iv) MAXIMIZAÇÃO

Sejam $P_j = (p_{j1},p_{j2},\dots,p_{jn})$, $j = 1,\dots,k$, $p_{ji} \geq 0$, $i=1,\dots,n$; $j=1,\dots,k$ e $\sum_{i=1}^{n} p_{ji} = 1$, $j = 1,\dots,k$; k distribuições discretas.

Se $P_1 = P_2 = \dots = P_k$, então a Afinidade entre estas k distribuições é máxima, ou seja, $\rho_{n,k} = 1$.

(1.1.6) $\rho_{n,k}(p_{11},\dots,p_{1n}:p_{21},\dots,p_{2n}:\dots:p_{k1},\dots,p_{kn}) \leq$

$\leq \rho_{n,k}(p_{11},\dots,p_{1n} : p_{11},\dots,p_{1n}:\dots: p_{11},\dots,p_{1n}) = 1.$

v) NÃO-NEGATIVIDADE

Sejam $P_j = (p_{j1},\ldots,p_{jn})$, $j = 1,\ldots,k$, tal que $p_{ji}=0$, $i=1,\ldots,n$; $j = 1,\ldots,k$ e $\sum_{i=1}^{n} p_{ji} = 1$, $j = 1,\ldots,k$, k distribuições discretas. Então:

$$(1.1.7) \quad \rho_{n,k}(p_{11},\ldots,p_{1n}:p_{21},\ldots,p_{2n}:\ldots:p_{k1},\ldots,p_{kn}) \geq 0.$$

vi) ESTRUTURA

Seja $P_j = (p_{j1},\ldots,p_{jn})$, $j = 1,\ldots,k$, tal que $p_{ji} \geq 0$, $i=1,\ldots,n$, $j= 1,\ldots,k$ e $\sum_{i=1}^{n} p_{ji} = 1$, $j = 1,\ldots,k$.

$$(1.1.8) \quad \rho_{n,k}(p_{11},\ldots,p_{1n}:p_{21},\ldots,p_{2n}:\ldots:p_{k1},\ldots,p_{kn}) =$$

$$= \sum_{i=1}^{n} f(p_{1i},p_{2i},\ldots,p_{ki}) \text{ , onde}$$

$$f(p_{1i},p_{2i},\ldots,p_{ki}) = [\prod_{j=1}^{k} p_{ji}]^{1/k} = \prod_{j=1}^{k} (p_{ji})^{1/k}$$

ou ainda:

$$(1.1.9) \quad \rho_{n,k}(p_{11},\ldots,p_{1n}:p_{21},\ldots,p_{2n}:\ldots:p_{k1},\ldots,p_{kn}) =$$

$$= \sum_{i=1}^{n} [\prod_{j=1}^{k} f_j(p_i)] \text{ ,}$$

onde $f_j(p_i) = (p_{ji})^{1/k}$.

vii) EXPANSIBILIDADE

Sejam $P_j = (p_{j1},\ldots,p_{jn},0)$, $j=1,\ldots,k$, tais que $\sum_{i=1}^{n} p_{ji} = 1$, $j=1,\ldots,k$ e $p_{ji} \geq 0$, $i = 1,\ldots,n$, $j = 1,\ldots,k$.

$$\text{(1.1.10)} \quad \rho_{n+1,k}(p_{11},\ldots,p_{1n},0:p_{21},\ldots p_{2n},0:\ldots:p_{k1},\ldots,p_{kn},0) =$$

$$= \rho_{n,k}(p_{11},\ldots,p_{1n}:p_{21},\ldots,p_{2n}:\ldots:p_{k1},\ldots,p_{kn}).$$

viii) NULIDADE

Sejam $P_j = (p_{j1},\ldots,p_{jn})$, $j=1,\ldots,k$, tais que $\sum_{i=1}^{n} p_{ji} = 1$, $j = 1,\ldots,k$ e $p_{ji} \geq 0$, $i = 1,\ldots,n$, $j = 1,\ldots,k$.

(1.1.11) $\rho_{n,k}(p_{11},\ldots,p_{1n}:p_{21},\ldots,p_{2n}:\ldots:p_{k1},\ldots,p_{kn}) = 0$, sempre que $\prod_{j=1}^{k} p_{ji} = 0$, $i = 1,\ldots,n$.

ix) CONTINUIDADE

Sejam $P_j = (p_{j1},\ldots,p_{jn})$, $j = 1,\ldots,k$, tais que $\sum_{i=1}^{n} p_{ji} = 1$, $j=1,\ldots,k$ e $p_{ji} \geq 0$, $i = 1,\ldots,n$, $j = 1,\ldots,k$.

A Afinidade entre as distribuições $P_1,\ldots,P_k$ é uma função contínua em suas kn variáveis.

x) ADITIVIDADE FORTE

Sejam $P_j^{(\ell)} = (p_{j1}^{(\ell)},\ldots,p_{jm_\ell}^{(\ell)})$, $j=1,\ldots,k$ e $\ell = 1,\ldots,n$, tais que $\sum_{i=1}^{m_\ell} P_{ji}^{(\ell)} = 1$, $j = 1,\dots,k$, $\ell = 1,\dots,n$ e $P_{ji}^{(\ell)} \geq 0$, $i = 1,\dots,m_\ell$, $j = 1,\dots,k$ e $\ell = 1,\dots,n$. E sejam

$P_j = (p_{j1},\dots,p_{jn})$, $j=1,\dots,k$, tais que $\sum_{\ell=1}^{n} p_{j\ell} = 1$, $j=1,\dots,k$ e $p_{j\ell} \geq 0$, $j = 1,\dots,k$, $\ell = 1,\dots,n$.

Então:

$$
\begin{aligned}
(1.1.12)\ \rho_{\sum_{\ell=1}^{n} m_\ell}(&p_{11}P_{11}^{(1)},\dots,p_{11}P_{1m_1}^{(1)},p_{12}P_{11}^{(2)},\dots,p_{12}P_{1m_2}^{(2)},\dots,\\
&p_{1n}P_{11}^{(n)},\dots,p_{1n}P_{1m_n}^{(n)} : p_{21}P_{21}^{(1)},\dots,p_{21}P_{2m_1}^{(1)},p_{22}P_{21}^{(2)},\dots,\\
&p_{22}P_{2m_2}^{(2)},\dots,p_{2n}P_{21}^{(n)},\dots,p_{2n}P_{2m_n}^{(n)}:\dots:p_{k1}P_{k1}^{(1)},\dots,\\
&p_{k1}P_{km_1}^{(1)},p_{k2}P_{k1}^{(2)},\dots,p_{k2}P_{km_2}^{(2)},\dots,p_{kn}P_{k1}^{(n)},\dots,p_{kn}P_{km_n}^{(n)})\\
= \sum_{\ell=1}^{n} \Big(\prod_{j=1}^{k} p_{j\ell}\Big)^{1/k} \cdot \rho_{m_\ell}(&P_{11}^{(\ell)},\dots,P_{1m_\ell}^{(\ell)} : P_{21}^{(\ell)},\dots,P_{2m_\ell}^{(\ell)} :\dots:\\
&P_{k1}^{(\ell)},\dots,P_{km_\ell}^{(\ell)}) .
\end{aligned}
$$

PROVA:

Lado esquerdo de (1.1.12) =

$$
\begin{aligned}
&= [p_{11}P_{11}^{(1)}p_{21}P_{21}^{(1)}\dots p_{k1}P_{k1}^{(1)}]^{1/k} +\dots+ [p_{11}P_{1m_1}^{(1)}\ p_{21}P_{2m_1}^{(1)}\ \dots\\
&p_{k1}P_{km_1}^{(1)}]^{1/k} + [p_{12}P_{11}^{(2)}p_{22}P_{21}^{(2)}\dots p_{k2}P_{k1}^{(2)}]^{1/k} + \dots +
\end{aligned}
$$

$$+ [p_{12}P^{(2)}_{1m_2}\, p_{22}P^{(2)}_{2m_2} \cdots p_{k2}P^{(2)}_{km_2}]^{1/k} + \ldots + [p_{1n}P^{(n)}_{11}\, p_{2n}P^{(n)}_{21} \cdots$$

$$p_{kn}P^{(n)}_{k1}]^{1/k} + \ldots + [p_{1n}P^{(n)}_{1m_n}\, p_{2n}P^{(n)}_{2m_n} \cdots p_{kn}P^{(n)}_{km_n}]^{1/k} =$$

$$= [\prod_{j=1}^{k} p_{j1}]^{1/k} \{\sum_{i=1}^{m_1} [\prod_{j=1}^{k} P^{(1)}_{ji}]^{1/k}\} + [\prod_{j=1}^{k} p_{j2}]^{1/k} \{\sum_{i=1}^{m_2} [\prod_{j=1}^{k} P^{(2)}_{ji}]^{1/k}\} +$$

$$+ \ldots + [\prod_{j=1}^{k} p_{jn}]^{1/k} \{\sum_{i=1}^{m_n} [\prod_{j=1}^{k} P^{(n)}_{ji}]^{1/k}\} =$$

$$= \sum_{\ell=1}^{n} [\prod_{j=1}^{k} p_{j\ell}]^{1/k} \cdot \{\sum_{i=1}^{m_\ell} [\prod_{j=1}^{k} P^{(\ell)}_{ji}]^{1/k}\}$$

$$= \sum_{\ell=1}^{n} [\prod_{j=1}^{k} p_{j\ell}]^{1/k} \cdot \rho_{m_\ell,k}(P^{(\ell)}_{11}, \ldots, P^{(\ell)}_{1m_\ell} : P^{(\ell)}_{21}, \ldots, P^{(\ell)}_{2m_\ell} :$$

$$: \ldots : P^{(\ell)}_{k1}, \ldots, P^{(\ell)}_{km_\ell}).$$

xi) PRODUTO

Sejam $P_j = (p_{j1}, \ldots, p_{jn})$ e $Q_j = (q_{j1}, \ldots, q_{jm})$, $j = 1,=,\ldots,k$, tais que

$$\sum_{i=1}^{n} p_{ji} = 1 = \sum_{\ell=1}^{m} q_{j\ell} \text{ , para } j = 1,\ldots,k \text{ ;}$$

$p_{ji} \geq 0$, $i = 1,\ldots,n$ e $q_{j\ell} \geq 0$, $\ell = 1,\ldots,m$, para $j=1,\ldots,k$.

Então:

$$(1.1.13)\quad \rho_{nm,k}(p_{11}q_{11},\ldots,p_{11}q_{1m},p_{12}q_{11},\ldots,p_{12}q_{1m},\ldots,$$

$$p_{1n}q_{11},\ldots,p_{1n}q_{1m} : p_{21}q_{21},\ldots,q_{21}q_{2m},p_{22}q_{21},\ldots,$$

$$p_{22}q_{2m},\ldots,p_{2n}q_{21},\ldots,q_{2n}q_{2m}:\ldots:p_{k1}q_{k1},\ldots,$$

$$p_{k1}q_{km},p_{k2}q_{k1},\ldots,p_{k2}q_{km},\ldots,p_{kn}q_{k1},\ldots,p_{kn}q_{km})=$$

$$= \rho_{n,k}(p_{11},\ldots,p_{1n}:p_{21},\ldots,p_{2n}:\ldots:p_{k1},\ldots,p_{kn}) \cdot$$

$$\cdot\rho_{m,k}(q_{11},\ldots,q_{1m}:q_{21},\ldots,q_{2m}:\ldots:q_{k1},\ldots,q_{km}).$$

PROVA:

Lado esquerdo de (1.1.13) =

$$= [\prod_{j=1}^{k} p_{j1}q_{j1}]^{1/k} +\ldots+ [\prod_{j=1}^{k} p_{j1}q_{jm}]^{1/k} +\ldots+ [\prod_{j=1}^{k} p_{j2}q_{j1}]^{1/k} +$$

$$\ldots+ [\prod_{j=1}^{k} p_{j2}q_{jm}]^{1/k} +\ldots+ [\prod_{j=1}^{k} p_{jn}q_{j1}]^{1/k} +\ldots+ [\prod_{j=1}^{k} p_{jn}q_{jm}]^{1/k} =$$

$$= [\prod_{j=1}^{k} p_{j1}]^{1/k} [\sum_{\ell=1}^{m} (\prod_{j=1}^{k} q_{j\ell})^{1/k}] + [\prod_{j=1}^{k} p_{j2}]^{1/k} [\sum_{\ell=1}^{m} (\prod_{j=1}^{k} q_{j\ell})^{1/k}] +\ldots+$$

$$+ [\prod_{j=1}^{k} p_{jn}]^{1/k} [\sum_{\ell=1}^{m} (\prod_{j=1}^{k} q_{j\ell})^{1/k}] = [\sum_{i=1}^{n} (\prod_{j=1}^{k} p_{ji})^{1/k}] \cdot [\sum_{\ell=1}^{m} (\prod_{j=1}^{k} q_{j\ell})^{1/k}] =$$

$$= \rho_{n,k}(p_{11},\ldots,p_{1n} : p_{21},\ldots,p_{2n}:\ldots: p_{k1},\ldots,p_{kn}) \cdot$$

$\cdot\ \rho_{m,k}(q_{11},\dots,q_{1m}:q_{21},\dots,q_{2m}:\dots:q_{k1},\dots,q_{km})$.

xii) NORMALIZAÇÃO

Se $P_i = (\frac{1}{2}, \frac{1}{2})$, $i = 1,2,\dots,k-2$; $P_{k-1} = (\frac{1}{4}, \frac{3}{4})$ e $P_k = (\frac{3}{4}, \frac{1}{4})$,

então:

$$(1.1.14)\quad \rho_{2,k}(\tfrac{1}{2}, \tfrac{1}{2} :\dots: \tfrac{1}{2}, \tfrac{1}{2} : \tfrac{1}{4}, \tfrac{3}{4} : \tfrac{3}{4}, \tfrac{1}{4}) = (\tfrac{3}{4})^{1/k} \;,\; k \geq 2 \;.$$

1.2 - AFINIDADE ENTRE DUAS POPULAÇÕES

1.2.1 - DEFINIÇÃO

Sejam $P = (p_1,\dots,p_n)$ e $Q = (q_1,\dots,q_n)$ duas distribuições discretas, ou seja, os p_i's e os q_i's são números tais que $\sum_{i=1}^{n} p_i = 1 = \sum_{i=1}^{n} q_i$ e $p_i \geq 0$, $q_i \geq 0$, $i = 1,\dots,n$.

A Afinidade entre as distribuições P e Q , denotada por $\rho_{n,2}$ é definida por:

$$(1.2.1)\quad \rho_{n,2} \equiv \rho_{n,2}(p_1,\dots,p_n : q_1,\dots,q_n) = \sum_{i=1}^{n} (p_i q_i)^{1/2}.$$

Vemos claramente que $\rho_{n,2}$ é caso espacial de $\rho_{n,k}$, para $k = 2$.

1.2.2 - PROPRIEDADES

Nesta secção enunciaremos as principais propriedades da Afinidade entre duas distribuições, sem as provas, pois são análogas às feitas para o caso em que temos várias populações.

i) RECURSIVIDADE

Sejam $P = (p_1,\ldots,p_n)$ e $Q = (q_1,\ldots,q_n)$ tais que $\sum_{i=1}^{n} p_i = 1 = \sum_{i=1}^{n} q_i$, $p_i \geq 0$, $q_i \geq 0$, $i = 1,\ldots,n$, duas distribuições discretas. Então:

$$(1.2.2)\quad \rho_{n,2}(p_1,\ldots,p_n:q_1,\ldots,q_n) = \rho_{n-1,2}(p_1 + p_2, p_3,\ldots,p_n : q_1 + q_2, q_3,\ldots,q_n) + K(p_1,p_2 : q_1,q_2)\cdot \rho_{2,2}\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2} : \frac{q_1}{q_1+q_2}, \frac{q_2}{q_1+q_2}\right),$$

$$\forall\ n > 2,\ p_1+p_2 > 0 \text{ e } q_1+q_2 > 0,\ \text{onde:}$$

$$(1.2.3)\quad K(p_1,p_2:q_1,q_2) = \frac{[(p_1q_1)^{1/2}+(p_2q_2)^{1/2}-(p_1+p_2)^{1/2}(q_1+q_2)^{1/2}]\,(p_1+p_2)^{1/2}(q_1+q_2)^{1/2}}{(p_1q_1)^{1/2} + (p_2q_2)^{1/2}}$$

Esta propriedade facilita bastante no cálculo da Afinidade, quando consideramos a união de dois eventos mutuamente exclusivos.

ii) SIMETRIA

A medida de Afinidade $\rho_{n,2}$ é simétrica em pares $\{p_i, q_i\}$, $i = 1,\ldots,n$. Ou seja:

$$(1.2.4)\quad \rho_{n,2}(p_1,\ldots,p_n:q_1,\ldots,q_n) = \rho_{n,2}(p_{a_1},\ldots,p_{a_n} : q_{a_1},\ldots,q_{a_n}),$$

onde $\{a_1,\ldots,a_n\}$ é uma permutação arbitrária de $\{1,\ldots,n\}$.

iii) SIMETRIA EM DISTRIBUIÇÃO

Sejam $P = (p_1,\ldots,p_n)$ e $Q = (q_1,\ldots,q_n)$ duas distribuições discretas, ou seja, $\sum_{i=1}^{n} p_i = 1 = \sum_{i=1}^{n} q_i$, $p_i \geq 0$, $q_i \geq 0$,$i=1,\ldots,n$.

Então:

(1.2.5) $\rho_{n,2}(p_1,\ldots,p_n:q_1,\ldots,q_n) = \rho_{n,2}(q_1,\ldots,q_n:p_1,\ldots,p_n)$.

Ou, em outra notação:

$$\rho_{n,2}(P:Q) = \rho_{n,2}(Q:P) \text{ , onde}$$

$$\rho_{n,2}(P:Q) = \rho_{n,2}(p_1,\ldots,p_n : q_1,\ldots,q_n) \text{ .}$$

iv) MAXIMIZAÇÃO

Sejam $P = (p_1,\ldots,p_n)$ e $Q = (q_1,\ldots,q_n)$ tais que $\sum_{i=1}^{n} p_i = 1 = \sum_{i=1}^{n} q_i$, $p_i \geq 0$, $q_i \geq 0$, $i = 1,\ldots,n$.

Se $P = Q$, então a Afinidade entre P e Q é máxima, ou seja, $\rho_{n,2} = 1$.

(1.2.6) $\rho_{n,2}(p_1,\ldots p_n:q_1,\ldots,q_n) \leq \rho_n(p_1,\ldots,p_n:p_1,\ldots,p_n) = 1$

v) NÃO-NEGATIVIDADE

Sejam $P = (p_1,\ldots,p_n)$ e $Q = (q_1,\ldots,q_n)$ tais que $\sum_{i=1}^{n} p_i = 1 = \sum_{i=1}^{n} q_i$, $p_i \geq 0$, $q_i \geq 0$, $i = 1,\ldots,n$.

(1.2.7) $\rho_{n,2}(p_1,\ldots,p_n : q_1,\ldots,q_n) \geq 0$.

vi) ESTRUTURA

Sejam $P = (p_1,\ldots,p_n)$ e $Q = (q_1,\ldots,q_n)$, tais que $\sum_{i=1}^{n} p_i = 1 = \sum_{i=1}^{n} q_i$, $p_i \geq 0$, $q_i \geq 0$, $i = 1,\ldots,n$.

$$(1.2.8)\quad \rho_{n,2}(p_1,\ldots,p_n : q_1,\ldots,q_n) = \sum_{i=1}^{n} f(p_i,q_i) \text{ , onde}$$

$$f(p_i,q_i) = (p_i q_i)^{1/2} .$$

Ou ainda:

$$(1.2.9)\quad \rho_n(p_1,\ldots,p_n : q_1,\ldots,q_n) = \sum_{i=1}^{n} f(p_i)f(q_i), \text{ onde } f(p) = p^{1/2}.$$

vii) EXPANSIBILIDADE

Sejam $P = (p_1,\ldots,p_n,0)$ e $Q = (q_1,\ldots,q_n,0)$ tais que $\sum_{i=1}^{n} p_i = 1 = \sum_{i=1}^{n} q_i$ e $p_i \geq 0$, $q_i \geq 0$, $i = 1,\ldots,n$.

$$(1.2.10)\quad \rho_{n+1}(p_1,\ldots,p_n,0 : q_1,\ldots,q_n,0) = \rho_{n,2}(p_1,\ldots,p_n : q_1,\ldots,q_n)$$

viii) NULIDADE

Sejam $P = (p_1,\ldots,p_n)$ e $Q = (q_1,\ldots,q_n)$, tais que $\sum_{i=1}^{n} p_i = 1 = \sum_{i=1}^{n} q_i$, $p_i \geq 0$, $q_i \geq 0$, $i = 1,\ldots,n$. Então:

$$(1.2.11)\quad \rho_{n,2}(p_1,\ldots,p_n : q_1,\ldots,q_n) = 0 \text{ , sempre que } p_i q_i = 0 \text{ ,}$$

$i = 1,\ldots,n$.

ix) CONTINUIDADE

Sejam $P = (p_1,\ldots,p_n)$ e $Q = (q_1,\ldots,q_n)$, tais que $\sum_{i=1}^{n} p_i = 1 = \sum_{i=1}^{n} q_i$, $p_i \geq 0$, $q_i \geq 0$, $i = 1,\ldots,n$.

A Afinidade entre as distribuições P e Q, $\rho_{n,2}(P : Q)$, é uma função contínua em cada uma das suas $2n$ variáveis.

x) ADITIVIDADE FORTE

Seja: $P = (p_1,\dots,p_n)$ e $Q = (q_1,\dots,q_n)$, tais que $\sum_{j=1}^{n} p_j = 1 = \sum_{j=1}^{n} q_j$, $p_j \geq 0$, $q_j \geq 0$, $j = 1,\dots,n$.

E seja:

$$P^{(j)} = (P_1^{(j)},\dots,P_{m_j}^{(j)}) \text{ e } Q^{(j)} = (Q_1^{(j)},\dots,Q_{m_j}^{(j)}),$$

$j = 1,\dots,n$, tais que $\sum_{i=1}^{m_j} P_i^{(j)} = 1 = \sum_{i=1}^{m_j} Q_i^{(j)}$, $j = 1,\dots,n$;

$P_i^{(j)} \geq 0$ e $Q_i^{(j)} \geq 0$, $i = 1,\dots,m_j$, $j = 1,\dots,n$.

Então:

$$(1.2.12)\quad \rho_{(m_1+\dots+m_n),2}(p_1P_1^{(1)},\dots,p_1P_{m_1}^{(1)},\dots,p_nP_1^{(n)},\dots,p_nP_{m_n}^{(n)}: q_1Q_1^{(1)},\dots,q_1Q_{m_1}^{(1)},\dots,q_nQ_1^{(n)},\dots,q_nQ_{m_n}^{(n)}) =$$

$$= \sum_{j=1}^{n} (p_jq_j)^{1/2} \cdot \rho_{(m_j),2}(P_1^{(j)},\dots,P_{m_j}^{(j)}: Q_1^{(j)},\dots,Q_{m_j}^{(j)}).$$

xi) PRODUTO

Sejam $P = (p_1,\dots,p_m)$; $Q = (q_1,\dots,q_n)$; $R = (r_1,\dots,r_m)$; $S = (s_1,\dots,s_n)$, tais que $\sum_{i=1}^{m} p_i = 1 = \sum_{i=1}^{m} r_i$, $\sum_{j=1}^{n} q_j = 1 = \sum_{j=1}^{n} s_j$; $p_i \geq 0$,

$r_i \geq 0$, $i = 1,\dots,m$; $q_j \geq 0$, $s_j \geq 0$, $j = 1,\dots,n$.

Então:

$$(1.2.13)\quad \rho_{mn,2}(p_1q_1,\ldots,p_1q_n,p_2q_1,\ldots,p_2q_n,\ldots,p_mq_1,\ldots,p_mq_n:$$

$$r_1s_1,\ldots,r_1s_n,\ r_2s_1,\ldots,r_2s_n,\ldots,\ r_ms_1,\ldots,r_ms_n) =$$

$$= \rho_{m,2}(p_1,\ldots,p_m : r_1,\ldots,r_m)\cdot\rho_{n,2}(q_1,\ldots,q_n:s_1,\ldots,s_n)$$

xii) NORMALIZAÇÃO

Se $P = (\frac{1}{4}, \frac{3}{4})$ e $Q = (\frac{3}{4}, \frac{1}{4})$, então

$$(1.2.14)\quad \rho_{2,2}(\tfrac{1}{4}, \tfrac{3}{4} : \tfrac{3}{4}, \tfrac{1}{4}) = (\tfrac{3}{4})^{1/2}\ .$$

CAPÍTULO II

FUNDAMENTO AXIOMÁTICO

Vimos no capítulo anterior, as definições e uma série de propriedades da Afinidade, tanto para o caso em que temos duas populações, como para o caso em que temos mais de duas. Agora, daremos uma caracterização axiomática para esta medida; ou seja, daremos uma definição axiomática, através de um teorema, baseado em três das propriedades, usando-as como postulados, em cada caso considerado.

Este teorema nos permitirá afirmar se uma função arbitrária, satisfazendo três particulares postulados, é a medida que estamos estudando. Aqui, tomaremos como postulados as propriedades de Recursividade, Simetria e Normalização. Mas, um outro conjunto de propriedades convenientes, pode ser tomado como um conjunto de postulados, para o teorema.

Na primeira secção deste capítulo, trataremos do caso em que temos várias distribuições e na segunda enfocaremos o caso em que temos somente duas.

2.1 - DEFINIÇÃO AXIOMÁTICA DE AFINIDADE ENTRE VÁRIAS DISTRIBUIÇÕES

Com a definição axiomática que daremos aqui, poderemos dizer se uma função arbitrária, satisfazendo um dado conjunto de postulados é unicamente determinada como a medida de Afinidade.

Daremos esta caracterização pelo teorema seguinte:

TEOREMA:

Sejam $P_j = (p_{j1},\dots,p_{jn})$, $j = 1,\dots,k$, com $p_{ji} \geq 0$, $i=1,\dots,n$;

$j = 1,\ldots,k$ e $\sum_{i=1}^{n} p_{ji} = 1$, $j = 1,\ldots,k$; k distribuições discretas.

Seja $\rho_{n,k}(p_{11},\ldots,p_{1n} : p_{21},\ldots,p_{2n} : \ldots : p_{k1},\ldots,p_{kn})$, para $n \geq 2$, uma função arbitrária dos p's.

Então, $\rho_{n,k}$ é unicamente determinada por:

$$\rho_{n,k} = \sum_{i=1}^{n} [\prod_{j=1}^{k} p_{ji}]^{1/k}$$

se satisfizer os seguintes postulados:

(1) RECURSIVIDADE

$$\rho_{n,k}(p_{11},\ldots,p_{1n} : p_{21},\ldots,p_{2n} :\ldots: p_{k1},\ldots,p_{kn}) =$$

$$= \rho_{n-1,k}(p_{11}+p_{12},p_{13},\ldots,p_{1n}:p_{21}+p_{22},p_{23},\ldots,p_{2n}:\ldots:p_{k1}+p_{k2},p_{k3},\ldots,$$

$$p_{kn}) + K(p_{11},p_{12}:p_{21},p_{22}:\ldots:p_{k1},p_{k2})\cdot\rho_{2,k}(\frac{p_{11}}{p_{11}+p_{12}}, \frac{p_{12}}{p_{11}+p_{12}} :$$

$$\frac{p_{21}}{p_{21}+p_{22}}, \frac{p_{22}}{p_{21}+p_{22}} :\ldots: \frac{p_{k1}}{p_{k1}+p_{k2}}, \frac{p_{k2}}{p_{k1}+p_{k2}}) ,$$ onde

$$K(p_{11},p_{12} : p_{21},p_{22}:\ldots:p_{k1},p_{k2}) =$$

$$= \frac{\{\sum_{i=1}^{2} [\prod_{j=1}^{k} p_{ji}]^{1/k} - [\prod_{j=1}^{k} (\sum_{i=1}^{2} p_{ji})]^{1/k}\} [\prod_{j=1}^{k} (\sum_{i=1}^{2} p_{ji})]^{1/k}}{\sum_{i=1}^{2} [\prod_{j=1}^{k} p_{ji}]^{1/k}}$$

$$\forall\ n \geq 2 , \quad \sum_{i=1}^{2} p_{ji} > 0 , \ j=1,\ldots,k$$

(2) SIMETRIA

$\rho_{3,k}$ é simétrica em k-uplas $\{p_{1i},\ldots,p_{ki}\}$, $i = 1,2,3$; ou seja:

$\rho_{3,k}(p_{11},p_{12},p_{13}:p_{21},p_{22},p_{23}:\ldots:p_{k1},p_{k2},p_{k3}) =$

$= \rho_{3,k}(p_{1a_1},p_{1a_2},p_{1a_3}:p_{2a_1},p_{2a_2},p_{2a_3}:\ldots:p_{ka_1},p_{ka_2},p_{ka_3})$, onde

$\{a_1, a_2, a_3\}$ é uma permutação arbitrária de $\{1,2,3\}$.

(3) NORMALIZAÇÃO

Seja: $P_i = (\frac{1}{2}, \frac{1}{2})$, $i = 1,2,\ldots,k-2$; $P_{k-1} = (\frac{1}{4}, \frac{3}{4})$ e $P_k(\frac{3}{4}, \frac{1}{4})$,

então:

$$\rho_{2,k}(\frac{1}{2}, \frac{1}{2} : \frac{1}{2}, \frac{1}{2} : \ldots : \frac{1}{2}, \frac{1}{2} : \frac{1}{4}, \frac{3}{4} : \frac{3}{4}, \frac{1}{4}) = (\frac{3}{4})^{1/k} \quad , \ k = 2,3,\ldots$$

PROVA

Para esta demonstração, vamos precisar de alguns resultados, obtidos pelos lemas seguintes:

LEMMA 2.1.1 - Seja

(2.1.1) $f(x_1,x_2,\ldots,x_k) = \rho_{2,k}(x_1,1-x_1 : x_2,1-x_2:\ldots:x_k,1-x_k)$,

para todo $x_j \in [0,1]$, $j = 1,2,\ldots,k$.

Então:

(2.1.2) $f(x_1,x_2,\ldots,x_k) = f(1-x_1, 1-x_2,\ldots,1-x_k)$ para todo

$x_j \in [0,1]$, $j = 1,2,\ldots,k$.

PROVA:

Usando o Postulado (2) , temos:

$$(2.1.3)\quad \rho_{3,k}(x_1,1-x_1-u_1,u_1:x_2,1-x_2-u_2,u_2:\dots:x_k,1-x_k-u_k,u_k) =$$

$$= \rho_{3,k}(1-x_1-u_1,x_1,u_1 : 1-x_2-u_2,x_2,u_2:\dots:1-x_k-u_k, x_k,u_k)$$

Pelo Postulado (1), a equação (2.1.3) pode ser escrita como:

$$(2.1.4)\quad \rho_{2,k}(1-u_1, u_1: 1-u_2,u_2:\dots:1-u_k,u_k) +$$

$$+ \frac{\{[\prod_{j=1}^{k} x_j]^{1/k} + [\prod_{j=1}^{k}(1-x_j-u_j)]^{1/k} - [\prod_{j=1}^{k}(1-u_j)]^{1/k}\}[\prod_{j=1}^{k}(1-u_j)]^{1/k}}{[\prod_{j=1}^{k} x_j]^{1/k} + [\prod_{j=1}^{k}(1-x_j-u_j)]^{1/k}} \cdot$$

$$\cdot\ \rho_{2,k}\left(\frac{x_1}{1-u_1}, \frac{1-x_1-u_1}{1-u_1} : \frac{x_2}{1-u_2}, \frac{1-x_2-u_2}{1-u_2}:\dots:\frac{x_k}{1-u_k}, \frac{1-x_k-u_k}{1-u_k}\right) =$$

$$= \rho_{2,k}(1-u_1, u_1 : 1-u_2, u_2:\dots: 1-u_k, u_k) +$$

$$+ \frac{\{[\prod_{j=1}^{k}(1-x_j-u_j)]^{1/k} + [\prod_{j=1}^{k} x_j]^{1/k} - [\prod_{j=1}^{k}(1-u_j)]^{1/k}\}[\prod_{j=1}^{k}(1-u_j)]^{1/k}}{[\prod_{j=1}^{k}(1-x_j-u_j)]^{1/k} + [\prod_{j=1}^{k} x_j]^{1/k}} \cdot$$

$$\cdot\ \rho_{2,k}\left(\frac{1-x_1-u_1}{1-u_1},\frac{x_1}{1-u_1}:\frac{1-x_2-u_2}{1-u_2},\frac{x_2}{1-u_2}:\ldots:\frac{1-x_k-u_k}{1-u_k},\frac{x_k}{1-u_k}\right)$$

Portanto

$$(2.1.5)\quad \rho_{2,k}\left(\frac{x_1}{1-u_1},\frac{1-x_1-u_1}{1-u_1}:\frac{x_2}{1-u_2},\frac{1-x_2-u_2}{1-u_2}:\ldots:\frac{x_k}{1-u_k},\frac{1-x_k-u_k}{1-u_k}\right) =$$

$$= \rho_{2,k}\left(\frac{1-x_1-u_1}{1-u_1},\frac{x_1}{1-u_1}:\frac{1-x_2-u_2}{1-u_2},\frac{x_2}{1-u_2}:\ldots:\frac{1-x_k-u_k}{1-u_k},\frac{x_k}{1-u_k}\right)$$

Usando (2.1.1) em (2.1.5), obtemos:

$$(2.1.6)\quad f\left(\frac{x_1}{1-u_1},\frac{x_2}{1-u_2},\ldots,\frac{x_k}{1-u_k}\right) = f\left(\frac{1-x_1-u_1}{1-u_1},\frac{1-x_2-u_2}{1-u_2},\ldots,\frac{1-x_k-u_k}{1-u_k}\right)$$

De onde obtemos (2.1.2).

Logo, o Lema 2.1.1. está provado.

LEMA 2.1.2: $f(x_1,x_2,\ldots,x_k)$ satisfaz a equação funcional:

$$(2.1.7)\quad f(x_1,\ldots,x_k)+\frac{\{[\prod\limits_{j=1}^{k} u_j]^{1/k}+[\prod\limits_{j=1}^{k}(1-x_j-u_j)]^{1/k}-[\prod\limits_{j=1}^{k}(1-x_j)]^{1/k}\}[\prod\limits_{j=1}^{k}(1-x_j)]^{1/k}}{[\prod\limits_{j=1}^{k} u_j]^{1/k}+[\prod\limits_{j=1}^{k}(1-x_j-u_j)]^{1/k}}$$

$$\cdot\ f\left(\frac{u_1}{1-x_1},\frac{u_2}{1-x_2},\ldots,\frac{u_k}{1-x_k}\right) = f(u_1,\ldots,u_k) +$$

$$+ \frac{\{[\prod_{j=1}^{k} x_j]^{1/k} + [\prod_{j=1}^{k} (1-x_j-u_j)]^{1/k} - [\prod_{j=1}^{k} (1-u_j)]^{1/k}\} [\prod_{j=1}^{k} (1-u_j)]^{1/k}}{[\prod_{j=1}^{k} x_j]^{1/k} + [\prod_{j=1}^{k} (1-x_j-u_j)]^{1/k}} \cdot$$

$$\cdot\ f(\frac{x_1}{1-u_1}, \frac{x_2}{1-u_2}, \ldots, \frac{x_k}{1-u_k}) \quad , \quad x_j, u_j \in [0,1]$$

e $x_j + u_j \in [0,1]$, $j = 1,=,\ldots,k$.

PROVA:

Pelo Postulado (2), temos:

(2.1.8) $\rho_{3,k}(x_1, 1-x_1-u_1, u_1 : x_2, 1-x_2-u_2, u_2 : \ldots : x_k, 1-x_k-u_k, u_k) =$

$$= \rho_{3,k}(u_1, 1-x_1-u_1, x_1 : u_2, 1-x_2-u_2, x_2 : \ldots : u_k, 1-x_k-u_k, x_k)$$

Aplicando o Postulado (1) em (2.1.8), obtemos

(2.1.9) $\rho_{2,k}(1-u_1, u_1 : 1-u_2, u_2 : \ldots : 1-u_k, u_k) +$

$$+ \frac{\{[\prod_{j=1}^{k} x_j]^{1/k} + [\prod_{j=1}^{k} (1-x_j-u_j)]^{1/k} - [\prod_{j=1}^{k} (1-u_j)]^{1/k}\} [\prod_{j=1}^{k} (1-u_j)]^{1/k}}{[\prod_{j=1}^{k} x_j]^{1/k} + [\prod_{j=1}^{k} (1-x_j-u_j)]^{1/k}} \cdot$$

$$\cdot\ \rho_{2,k}(\frac{x_1}{1-u_1}, \frac{1-x_1-u_1}{1-u_1} : \frac{x_2}{1-u_2}, \frac{1-x_2-u_2}{1-u_2} : \ldots : \frac{x_k}{1-u_k}, \frac{1-x_k-u_k}{1-u_k}) =$$

$$= \rho_{2,k}(1-x_1,\ x_1 : 1-x_2,\ x_2 : \ldots : 1-x_k,\ x_k) +$$

$$+ \frac{\{[\prod_{j=1}^{k} u_j]^{1/k} + [\prod_{j=1}^{k} (1-x_j-u_j)]^{1/k} - [\prod_{j=1}^{k} (1-x_j)]^{1/k}\}[\prod_{j=1}^{k} (1-x_j)]^{1/k}}{[\prod_{j=1}^{k} u_j]^{1/k} + [\prod_{j=1}^{k} (1-x_j-u_j)]^{1/k}} \, \cdot$$

$$\cdot\ \rho_{2,k}\left(\frac{u_1}{1-x_1},\ \frac{1-x_1-u_1}{1-x_1} : \frac{u_2}{1-x_2},\ \frac{1-x_2-u_2}{1-x_2} : \ldots : \frac{u_k}{1-x_k},\ \frac{1-x_k-u_k}{1-x_k}\right)$$

Usando (2.1.2) em (2.1.9) , obtemos (2.1.7)

Logo o Lema 2.1.2. está provado.

LEMA 2.1.3

$$(2.1.10) \quad f(x_1,x_2,\ldots,x_k) = [\prod_{j=1}^{k} x_j]^{1/k} + [\prod_{j=1}^{k} (1-x_j)]^{1/k} \quad ,$$

$$x_j \in [0,1] \quad , \quad j = 1,\ldots,k.$$

PROVA:

Fazendo $u_j = 0$, $j = 1,=,\ldots,k$, em (2.1.7), obtemos:

$$(2.1.11) \quad f(x_1,x_2,\ldots,x_k) + 0 = f(0,0,\ldots,0) +$$

$$+ \frac{[\prod_{j=1}^{k} x_j]^{1/k} + [\prod_{j=1}^{k} (1-x_j)]^{1/k} - 1}{[\prod_{j=1}^{k} x_j]^{1/k} + [\prod_{j=1}^{k} (1-x_j)]^{1/k}} \cdot f(x_1,x_2,\ldots,x_k).$$

Substituindo (2.1.10) em (2.1.14), obtemos

(2.1.15) $\rho_{n,k}(p_{11},\ldots,p_{1n} : p_{21},\ldots,p_{2n} : \ldots : p_{k1},\ldots,p_{kn}) =$

$$= [\prod_{j=1}^{k} p_{jn}]^{1/k} + [\prod_{j=1}^{k} (1-p_{jn})]^{1/k} +$$

$$+ \sum_{i=2}^{n-1} \frac{\{[\prod_{j=1}^{k} r_{j(i-1)}]^{1/k} + [\prod_{j=1}^{k} p_{ji}]^{1/k} - [\prod_{j=1}^{k} r_{ji}]^{1/k}\}[\prod_{j=1}^{k} r_{ji}]^{1/k}}{[\prod_{j=1}^{k} r_{j(i-1)}]^{1/k} + [\prod_{j=1}^{k} p_{ji}]^{1/k}} \cdot$$

$$\cdot \{[\prod_{j=1}^{k} (\frac{p_{ji}}{r_{ji}})]^{1/k} + [\prod_{j=1}^{k} (1 - \frac{p_{ji}}{r_{ji}})]^{1/k}\} =$$

$$= [\prod_{j=1}^{k} p_{jn}]^{1/k} + [\prod_{j=1}^{k} r_{j(n-1)}]^{1/k} + \sum_{i=2}^{n-1} [\prod_{j=1}^{k} r_{j(i-1)}]^{1/k} +$$

$$+ \sum_{i=2}^{n-1} [\prod_{j=1}^{k} p_{ji}]^{1/k} - \sum_{i=2}^{n-1} [\prod_{j=1}^{k} r_{ji}]^{1/k} =$$

$$= \sum_{i=2}^{n} [\prod_{j=1}^{k} p_{ji}]^{1/k} + \sum_{i=1}^{n-1} [\prod_{j=1}^{k} r_{ji}]^{1/k} - \sum_{i=2}^{n-1} [\prod_{j=1}^{k} r_{ji}]^{1/k} =$$

$$= \sum_{i=2}^{n} [\prod_{j=1}^{k} p_{ji}]^{1/k} + [\prod_{j=1}^{k} p_{j1}]^{1/k} = \sum_{i=1}^{n} [\prod_{j=1}^{k} p_{ji}]^{1/k}$$

Portanto:

$$(2.1.12)\quad f(0,0,\ldots,0) = \frac{1}{\left[\prod_{j=1}^{k} x_j\right]^{1/k} + \left[\prod_{j=1}^{k}(1-x_j)\right]^{1/k}} \cdot f(x_1,x_2,\ldots,x_k)$$

Fazendo $x_1 = x_2 = \ldots = x_{k-2} = \frac{1}{2}$; $x_{k-1} = \frac{1}{4}$ e $x_k = \frac{3}{4}$, em (2.1.12), pelo Postulado (3), temos:

$$(2.1.13)\qquad f(0,0,\ldots,0) = 1.$$

De (2.1.12) e (2.1.13), obtemos (2.1.10).

Logo, o Lema 2.1.3. está provado.

CONTINUAÇÃO DA PROVA DO TEOREMA:

Usando o Postulado (1), sucessivamente, e fazendo $r_{ji} = \sum_{\ell=1}^{i} p_{j\ell}$, $j = 1,2,\ldots,k$, temos:

$$(2.1.14)\quad \rho_{n,k}(p_{11},\ldots,p_{1n} : p_{21},\ldots,p_{2n} : \ldots : p_{k1},\ldots,p_{kn}) =$$

$$= f(p_{1n}, p_{2n},\ldots,p_{kn}) +$$

$$+ \sum_{i=2}^{n-1} \frac{\left\{\left[\prod_{j=1}^{k} r_{j(i-1)}\right]^{1/k} + \left[\prod_{j=1}^{k} p_{ji}\right]^{1/k} - \left[\prod_{j=1}^{k} r_{ji}\right]^{1/k}\right\}\left[\prod_{j=1}^{k} r_{ji}\right]^{1/k}}{\left[\prod_{j=1}^{k} r_{j(i-1)}\right]^{1/k} + \left[\prod_{j=1}^{k} p_{ji}\right]^{1/k}} \cdot$$

$$\cdot\, f\left(\frac{p_{1i}}{r_{1i}}, \frac{p_{2i}}{r_{2i}}, \ldots, \frac{p_{ki}}{r_{ki}}\right)$$

Logo,

$$\rho_{n,k}(p_{11},\dots,p_{1n}:p_{21},\dots,p_{2n}:\dots:p_{k1},\dots,p_{kn}) = \sum_{i=1}^{n} [\prod_{j=1}^{k} p_{ji}]^{1/k}$$

O que prova o teorema.

2.2. DEFINIÇÃO AXIOMÁTICA DE AFINIDADE ENTRE DUAS DISTRIBUIÇÕES.

Nesta secção daremos uma definição axiomática para a medida de Afinidade entre duas distribuições, sem demonstrarmos o teorema, pois este é um caso especial do teorema da secção anterior.

TEOREMA:

Sejam $P = (p_1,\dots,p_n)$ e $Q = (q_1,\dots,q_n)$, tais que $p_i \geq 0$, $q_i \geq 0$, $i = 1,\dots,n$ e $\sum_{i=1}^{n} p_i = 1 = \sum_{i=1}^{n} q_i$, duas distribuições discretas.

Seja $\rho_{n,2}(p_1,\dots,p_n : q_1,\dots,q_n)$, para $n \geq 2$, uma função arbitrária dos p's e dos q's.

Então, $\rho_{n,2}$ é unicamente determinada por:

$$\rho_{n,2}(p_1,\dots,p_n : q_1,\dots,q_n) = \sum_{i=1}^{n} (p_i q_i)^{1/2} ,$$

se satisfizer os seguintes postulados:

(1) RECURSIVIDADE

$$\rho_{n,2}(p_1,\dots,p_n : q_1,\dots,q_n) = \rho_{n-1,2}(p_1+p_2,p_3,\dots,p_n : q_1+q_2,q_3,\dots q_n)+$$

$$+\frac{[(p_1q_1)^{1/2}+(p_2q_2)^{1/2}-(p_1+p_2)^{1/2}(q_1+q_2)^{1/2}]\,(p_1+p_2)^{1/2}(q_1+q_2)^{1/2}}{(p_1q_1)^{1/2}+(p_2q_2)^{1/2}}\,.$$

$$\cdot\ \rho_{2,2}\left(\frac{p_1}{p_1+p_2}\,,\ \frac{p_2}{p_1+p_2}\ :\ \frac{q_1}{q_1+q_2}\,,\ \frac{q_2}{q_1+q_2}\right)\quad \text{para}\quad n \geq 2\ ,$$

$$p_1 + p_2 > 0 \quad , \quad q_1 + q_2 > 0\ .$$

(2) SIMETRIA

$\rho_{3,2}$ é simétrica em pares $\{p_i,q_i\}$, $i = 1,2,3$; ou seja ;

$$\rho_{3,2}(p_1, p_2, p_3 : q_1, q_2, q_3) = \rho_{3,2}(p_{a_1}, p_{a_2}, p_{a_3} : q_{a_1}, q_{a_2}, q_{a_3})\ ,$$

onde $\{a_1, a_2, a_3\}$ é uma permutação arbitrária de $\{1,2,3\}$.

(3) NORMALIZAÇÃO

$$\rho_{2,2}\left(\tfrac{1}{4}, \tfrac{3}{4} : \tfrac{3}{4}, \tfrac{1}{4}\right) = \left(\tfrac{3}{4}\right)^{1/2}\ .$$

CAPÍTULO III

APLICAÇÃO

O objetivo deste capítulo é o de ilustrar, com um exemplo prático, uma das possíveis aplicações da Medida de Afinidade.

Através deste exemplo e de outros apresentados no Capítulo IV, veremos como utilizar a Afinidade para mostrar a "proximidade" de populações, ou o comportamento de grupos, quanto a uma determinada característica, ao longo do tempo.

Estudaremos, aqui, o comportamento de 5 grupos classificados como: Residencial, Comercial, Industrial e Demais Classes, quanto ao consumo de energia elétrica no período de 1970 a 1978.

3.1 - DADOS E MEDIDAS DE AFINIDADE

Os dados utilizados foram obtidos de uma publicação da Coordenadoria de Análise de Dados (1978), atualmente Fundação SEADE.

Os mesmos se referem ao consumo de energia elétrica por classe, nas diversas Regiões Administrativas do Estado de São Paulo, no período de 1970 a 1978.

Para alcançar o nosso objetivo, o da análise, devemos ter os dados numa sequência de anos, separadamente, o que possibilita melhor estudo da tendência. Assim sendo, obtivemos um conjunto de tabelas, sendo duas para cada ano estudado.

A primeira tabela, para um dado ano, nos dá o consumo de energia elétrica por classe, em valor absoluto, e entre parênteses, as proporções de consumo em cada região para uma dada classe.

Na segunda tabela, obtemos uma matriz da medida de Afinidade, sempre calculada entre duas classes. Por exemplo, considerando o ano de 1970, a Afinidade entre as classes Residencial e Comercial é igual a 0,99565; e entre Residencial e Industrial é 0,98968; e assim por diante, como pode ser visto nas tabelas de números I a XVIII, referentes aos anos de 1970 a 1978, respectivamente.

TABELA I - ESTADO DE SÃO PAULO - REGIÕES ADMINISTRATIVAS - CONSUMO DE ENERGIA ELÉTRICA POR CLASSE - 1970 .

REGIÕES	RES.	COMERC.	INDUSTR.	RURAL	DEMAIS CLASSES (1)
Grande São Paulo	2217321. (0.6427)	1496340. (0.6987)	5339070. (0.6131)	4569. (0.0242)	914030. (0.5158)
Litoral	223896. 0.0649	179990. 0.0840	729291. 0.0837	3992. 0.0211	93746. 0.9529
Vale do Paraiba	98091. 0.0234	58009. 0.0271	338937. 0.0399	3560. 0.0188	59436. 0.0335
Sorocaba	108962. 0.0316	47665. 0.0223	515516. 0.0592	10318. 0.0546	77803. 0.0439
Campinas	330555. 0.0958	145552. 0.0680	1102357. 0.1266	75052. 0.3973	194961. 0.1100
Ribeirão Preto	179187. 0.0519	75290. 0.0352	248295. 0.0285	35008. 0.1853	147207. 0.0831
Bauru	67974. 0.0197	29121. 0.0136	119078. 0.0137	16042. 0.0849	59276. 0.0335
S.José R. Preto	76236. 0.0221	36253. 0.0169	48356. 0.0056	9452. 0.0500	61380. 0.0346
Araçatuba	45665. 0.0132	20290. 0.0095	130931. 0.0150	9366. 0.0496	35991. 0.0203
Presidente Prudente	45829. 0.0133	24918. 0.0116	56262. 0.0065	6680. 0.0354	48836. 0.0276
Marília	56274. 0.0163	28058. 0.0131	80027. 0.0092	14875. 0.0787	79410. 0.0448
Total do Estado	3449993.	2141486.	8708120.	188914.	1772076.

FONTE: CESP

(1) Incluem: Iluminação Pública, Poderes Públicos, Empresas de Serviços Públicos Interdepartamental e Tração Elétrica.

TABELA II

MEDIDA DE AFINIDADE
CONSUMO DE ENERGIA ELETRICA EM SAO PAULO

ANO DE 1970

TIPO	RESIDENCIAL	COMERCIAL	INDUSTRIAL	RURAL	DEMAIS CASOS
RESIDENCIAL	1.00000	.99565	.98968	.67688	.98673
COMERCIAL	.99565	1.00000	.98653	.61180	.97049
INDUSTRIAL	.98968	.98653	1.00000	.66480	.96908
RURAL	.67688	.61180	.66480	1.00000	.76965
DEMAIS CASOS	.98673	.97049	.96908	.76965	1.00000

TABELA III - ESTADO DE SÃO PAULO - REGIÕES ADMINISTRATIVAS - CONSUMO DE ENERGIA ELÉTRICA POR CLASSE - 1971 .

REGIÕES	RES.	COMERC.	INDUSTR.	RURAL	DEMAIS CLASSES (1)
Grande São Paulo	2465106. 0.6475	1700365. 0.7092	6085757. 0.6001	4468. 0.0203	963563. 0.4975
Litoral	247740. 0.0651	188792. 0.0787	830081. 0.0819	4535. 0.0206	95307. 0.0492
Vale do Paraiba	111285. 0.0292	58177. 0.0243	406708. 0.0401	5017. 0.0228	68244. 0.0352
Sorocaba	116885. 0.0307	53943. 0.0225	761247. 0.0751	14899. 0.0678	119054. 0.0615
Campinas	359109. 0.0943	162095. 0.0676	1314726. 0.1296	86340. 0.3931	219961. 0.1136
Ribeirão Preto	192726. 0.0506	81578. 0.0340	274118. 0.0270	38809. 0.1767	166633. 0.0860
Bauru	72914. 0.0192	31892. 0.0133	130560. 0.0129	18054. 0.0822	65397. 0.0338
S.José R. Preto	81418. 0.0214	40554. 0.0169	52267. 0.0052	10766. 0.0490	68431. 0.0353
Araçatuba	50358. 0.0132	22290. 0.0093	148040. 0.0146	10578. 0.0482	38466. 0.0199
Presidente Prudente	49522. 0.0130	26829. 0.0112	58828. 0.0058	8419. 0.0383	50299. 0.0260
Marília	60239. 0.0158	30909. 0.0129	78561. 0.0077	17755. 0.0808	81632. 0.0421
Total do Estado	3807303.	2397424.	10140893.	219640.	1936987.

FONTE: CESP

(1) Incluem: Iluminação Pública, Poderes Públicos, Empresas de Serviços Públicos Interdepartamental e Tração Elétrica.

TABELA IV

MEDIDA DE AFINIDADE

CONSUMO DE ENERGIA ELETRICA EM SAO PAULO

ANO DE 1971

TIPO	RESIDENCIAL	COMERCIAL	INDUSTRIAL	RURAL	DEMAIS CASOS
RESIDENCIAL	1.00000	.99593	.98600	.66540	.98244
COMERCIAL	.99593	1.00000	.98107	.59963	.96397
INDUSTRIAL	.98600	.98107	1.00000	.66296	.96663
RURAL	.66540	.59963	.66296	1.00000	.77511
DEMAIS CASOS	.98244	.96397	.96663	.77511	1.00000

TABELA V - ESTADO DE SÃO PAULO - REGIÕES ADMINISTRATIVAS - CONSUMO DE ENERGIA ELÉTRICA POR CLASSE - 1972 .

REGIÕES	RES.	COMERC.	INDUSTR.	RURAL	DEMAIS CLASSES (1)
Grande São Paulo	2679275. 0.6510	1810869. 0.7072	6865339. 0.5994	7024. 0.0275	1002590. 0.4541
Litoral	262804. 0.0639	184440. 0.0720	970289. 0.0847	5358. 0.0209	121955. 0.0552
Vale do Paraiba	122427. 0.0297	61412. 0.0240	503200. 0.0439	8101. 0.0317	76170. 0.0345
Sorocaba	125959. 0.0306	65472. 0.0256	824661. 0.0720	20699. 0.0809	117567. 0.0533
Campinas	387399. 0.0942	179847. 0.0702	1477634. 0.1290	98592. 0.3855	391294. 0.1772
Ribeirão Preto	206299. 0.0501	90222. 0.0352	313418. 0.0274	43550. 0.1703	180257. 0.0817
Bauru	76824. 0.0187	34746. 0.0136	144078. 0.0126	19259. 0.0753	70281. 0.0318
S.José R. Preto	85977. 0.0209	44835. 0.0175	65525. 0.0057	12740. 0.0498	75229. 0.0341
Araçatuba	52403. 0.0127	23804. 0.0093	153410. 0.0134	11787. 0.0461	41246. 0.0187
Presidente Prudente	53227. 0.0129	29143. 0.0114	56014. 0.0049	9823. 0.0384	49444. 0.0224
Marília	63150. 0.0153	35754. 0.0140	79304. 0.0069	18824. 0.0736	81586. 0.0370
Total do Estado	4115744.	2560544.	11452871.	255757.	2207619.

FONTE: CESP

(1) Incluem: Iluminação Pública, Poderes Públicos, Empresas de Serviços Públicos Interdepartamental e Tração Pública.

TABELA VI

MEDIDA DE AFINIDADE
CONSUMO DE ENERGIA ELETRICA EM SAO PAULO

ANO DE 1972

TIPO	RESIDENCIAL	COMERCIAL	INDUSTRIAL	RURAL	DEMAIS CASOS
RESIDENCIAL	1.00000	.99705	.98615	.68349	.97598
COMERCIAL	.99705	1.00000	.98237	.62844	.95777
INDUSTRIAL	.98615	.98237	1.00000	.68414	.96579
RURAL	.68349	.62844	.68414	1.00000	.82471
DEMAIS CASOS	.97598	.95777	.96579	.82471	1.00000

TABELA VII - ESTADO DE SÃO PAULO . REGIÕES ADMINISTRATIVAS - CONSUMO DE ENERGIA ELÉTRICA POR CLASSE - 1973 .

REGIÕES	RES.	COMERC.	INDUSTR.	RURAL	DEMAIS CLASSES (1)
Grande São Paulo	2900628. 0.6466	2009800. 0.7014	7882868. 0.5950	10276. 0.0343	1111669. 0.4569
Litoral	285742. 0.0637	207089. 0.0723	1181583. 0.0892	5720. 0.0191	129659. 0.0533
Vale do Paraiba	136312. 0.0304	71123. 0.0248	649445. 0.0490	11098. 0.0370	85328. 0.0351
Sorocaba	139378. 0.0311	74201. 0.0259	861065. 0.0650	26327. 0.0878	137565. 0.0565
Campinas	429938. 0.0958	205366. 0.0717	1767862. 0.1334	112976. 0.3768	421242. 0.1731
Ribeirão Preto	229729. 0.0512	105062. 0.0367	379640. 0.0287	49447. 0.1649	200867. 0.0825
Bauru	83941. 0.0185	40200. 0.0140	169172. 0.0128	21369. 0.0713	76327. 0.0314
S. José do R. Preto	95633. 0.0213	51263. 0.0179	69403. 0.0052	15030. 0.0501	82644. 0.0340
Araçatuba	57573. 0.0128	26815. 0.0094	127020. 0.0096	14184. 0.0473	45960. 0.0189
Presidente Prudente	59949. 0.0134	33407. 0.0117	86846. 0.0066	12192. 0.0407	53592. 0.0220
Marília	68019. 0.0152	41163. 0.0144	73626. 0.0056	21183. 0.0707	88470. 0.0364
Total do Estado	4485932.	2865489.	13248530.	299802.	2433323.

FONTE: CESP

(1) Incluem: Iluminação Pública, Poderes Públicos, Empresas de Serviço Públicos Interdepartamental e Tração Pública.

TABELA VIII

MEDIDA DE AFINIDADE
CONSUMO DE ENERGIA ELETRICA EM SAO PAULO

ANO DE 1973

TIPO	RESIDENCIAL	COMERCIAL	INDUSTRIAL	RURAL	DEMAIS CASOS
RESIDENCIAL	1.00000	.99717	.98614	.70115	.97736
COMERCIAL	.99717	1.00000	.99261	.64853	.96008
INDUSTRIAL	.98614	.98261	1.00000	.69901	.96604
RURAL	.70115	.64853	.69901	1.00000	.83466
DEMAIS CASOS	.97736	.96008	.96604	.83466	1.00000

TABELA IX - ESTADO DE SÃO PAULO - REGIÕES ADMINISTRATIVAS - CONSUMO DE ENERGIA ELÉTRICA POR CLASSE - 1974 .

REGIÕES	RES.	COMERC.	INDUST.	RURAL	DEMAIS CLASSES (1)
Grande São Paulo	3196628. 0.6448	2203493. 0.6955	8839905. 0.6025	11696. 0.0339	1225749. 0.4466
Litoral	308754. 0.0623	234524. 0.0740	1320513. 0.0900	5922. 0.0172	144843. 0.0528
Vale do Paraíba	154173. 0.0311	83774. 0.0264	767254. 0.0523	11508. 0.0333	99477. 0.0362
Sorocaba	156311. 0.0315	81965. 0.0259	854818. 0.0583	29533. 0.0856	141861. 0.0517
Campinas	480779. 0.0970	227434. 0.0718	2037708. 0.1389	133372. 0.3865	473958. 0.1727
Ribeirão Preto	256338. 0.0517	120483. 0.0380	397829. 0.0271	58779. 0.1703	215166. 0.0784
Bauru	90482. 0.0183	45223. 0.0143	175558. 0.0120	24131. 0.0699	86670. 0.0316
São José R. Preto	109020. 0.0220	55955. 0.0177	74291. 0.0051	17305. 0.0501	96395. 0.0351
Araçatuba	63035. 0.0127	29177. 0.0092	57226. 0.0039	15450. 0.0448	89074. 0.0325
Presidente Prudente	66418. 0.0134	36580. 0.0115	72388. 0.0049	13895. 0.0403	79111. 0.0288
Marília	75849. 0.0153	49661. 0.0157	75391. 0.0051	23521. 0.0682	92242. 0.0336
Total do Estado	4957784.	3168269.	14672881.	345112.	2744546.

FONTE: CESP

(1) Incluem: Iluminação Pública, Poderes Públicos, Empresas de Serviço Públicos Interdepartamental e Tração Pública.

TABELA X

MEDIDA DE AFINIDADE
CONSUMO DE ENERGIA ELETRICA EM SAO PAULO

ANO DE 1974

TIPO	RESIDENCIAL	COMERCIAL	INDUSTRIAL	RURAL	DEMAIS CASOS
RESIDENCIAL	1.00000	.99715	.98416	.70039	.97541
COMERCIAL	.99715	1.00000	.98193	.64885	.95792
INDUSTRIAL	.93416	.98193	1.00000	.68506	.95610
RURAL	.70039	.64885	.68506	1.00000	.83729
DEMAIS CASOS	.97541	.95792	.95610	.83729	1.00000

TABELA XI - ESTADO DE SÃO PAULO - REGIÕES ADMINISTRATIVAS -
CONSUMO DE ENERGIA ELÉTRICA POR CLASSE . 1975 .

REGIÕES	RES.	COMERC.	INDUSTR.	RURAL	DEMAIS CLASSES (1)
Grande São Paulo	3491461. 0.6434	2399246. 0.6924	9265780. 0.5993	12586. 0.0324	1395750. 0.4684
Litoral	337252. 0.0622	253780. 0.0732	1534086. 0.0992	6495. 0.0167	150762. 0.0506
Vale do Paraiba	174173. 0.0321	100162. 0.0289	860649. 0.0557	12707. 0.0327	105475. 0.0354
Sorocaba	172167. 0.0317	91926. 0.0265	628479. 0.0407	34181. 0.0881	149822. 0.0503
Campinas	530287. 0.0977	251877. 0.0727	2222978. 0.1438	147250. 0.3794	455505. 0.1529
Ribeirão Preto	277061. 0.0511	132151. 0.0381	438206. 0.0283	69383. 0.1788	239817. 0.0805
Bauru	99577. 0.0184	50735. 0.0146	203239. 0.0131	26780. 0.0690	92112. 0.0309
S. José R. Preto	122786. 0.0226	63083. 0.0182	80730. 0.0052	20502. 0.0528	131653. 0.0442
Araçatuba	66927. 0.0123	32555. 0.0094	60436. 0.0039	17304. 0.0446	81924. 0.0275
Presidente Prudente	73926. 0.0136	39979. 0.0115	84894. 0.0055	15942. 0.0411	84253. 0.0283
Marília	80791. 0.0149	49423. 0.0143	80376. 0.0052	24975. 0.0644	92558. 0.0311
Total do Estado	5426416.	3464917.	15459853.	388105.	2979631.

FONTE: CESP

(1) Incluem: Iluminação Pública, Poderes Públicos, Empresas de Serviços Públicos Interdepartamental e Tração Pública.

TABELA XII

MEDIDA DE AFINIDADE
CONSUMO DE ENERGIA ELETRICA EM SAO PAULO

ANO DE 1975

TIPO	RESIDENCIAL	COMERCIAL	INDUSTRIAL	RURAL	DEMAIS CASOS
RESIDENCIAL	1.00000	.99740	.98508	.69829	.98004
COMERCIAL	.99740	1.00000	.98380	.64794	.96456
INDUSTRIAL	.98508	.98380	1.00000	.68069	.95723
RURAL	.69829	.64794	.68069	1.00000	.82203
DEMAIS CASOS	.93004	.96456	.95723	.82203	1.00000

TABELA XIII - ESTADO DE SÃO PAULO - REGIÕES ADMINISTRATIVAS - CONSUMO DE ENERGIA ELÉTRICA POR CLASSE - 1976 .

REGIÕES	RES.	COMERC.	INDUSTR.	RURAL	DEMAIS CLASSES (1)
Grande São Paulo	3855946. 0.6393	2539395. 0.6816	10166041. 0.5821	14236. 0.0341	1587119. 0.4860
Litoral	372818. 0.0618	261661. 0.0702	1748528. 0.1001	7113. 0.0171	165929. 0.0508
Vale do Paraiba	195982. 0.0325	111664. 0.0300	1051658. 0.0602	12598. 0.0302	114052. 0.0349
Sorocaba	193673. 0.0321	102962. 0.0276	740755. 0.0424	38643. 0.0927	162723. 0.0498
Campinas	601335. 0.0997	287929. 0.0773	2643092. 0.1513	151131. 0.3625	487924. 0.1494
Ribeirão Preto	316250. 0.0524	154313. 0.0414	529829. 0.0303	76254. 0.1829	251493. 0.0770
Bauru	112103. 0.0196	58467. 0.0157	244626. 0.0140	30231. 0.0725	96895. 0.0297
S.José R. Preto	140135. 0.0232	74328. 0.0200	93916. 0.0054	22200. 0.0533	149355. 0.0457
Araçatuba	71921. 0.0119	36605. 0.0098	70001. 0.0040	18427. 0.0442	78201. 0.0239
Presidente Prudente	90334. 0.0134	43912. 0.0118	87709. 0.0050	19637. 0.0447	73635. 0.0225
Marília	90484. 0.0150	54146. 0.0145	89119. 0.0051	27400. 0.0657	98401. 0.0301
Total do Estado	6031484.	3725382.	17465274.	416870.	3265727.

FONTE: CESP

(1) Incluem: Iluminação Pública, Poderes Públicos, Empresas de Serviços Públicos Interdepartamental e Tração Elétrica.

TABELA XIV

MEDIDA DE AFINIDADE
CONSUMO DE ENERGIA ELETRICA EM SAO PAULO

ANO DE 1976

TIPO	RESIDENCIAL	COMERCIAL	INDUSTRIAL	RURAL	DEMAIS CASOS
RESIDENCIAL	1.00000	.99818	.98371	.70491	.98435
COMERCIAL	.99818	1.00000	.98184	.66338	.97306
INDUSTRIAL	.98371	.98184	1.00000	.69180	.96262
RURAL	.70491	.66338	.69180	1.00000	.81467
DEMAIS CASOS	.98435	.97306	.96262	.81467	1.00000

TABELA XV - ESTADO DE SÃO PAULO - REGIÕES ADMINISTRATIVAS -
CONSUMO DE ENERGIA ELÉTRICA POR CLASSE - 1977 .

REGIÕES	RES.	COMERC.	INDUSTR.	RURAL	DEMAIS CLASSES (1)
Grande São Paulo	4399663. 0.6388	2486801. 0.6650	10725455. 0.5591	16566. 0.0345	1749491. 0.5015
Litoral	435138. 0.0632	258515. 0.0691	1969163. 0.1026	7839. 0.0163	182360. 0.0523
Vale do Paraiba	227163. 0.0330	109366. 0.0292	1168503. 0.0609	16191. 0.0338	123070. 0.0353
Sorocaba	223380. 0.0324	107353. 0.0287	1056872. 0.0551	45716. 0.0953	167663. 0.0481
Campinas	685040. 0.0995	310469. 0.0830	3009985. 0.1569	166503. 0.3472	505069. 0.1448
Ribeirão Preto	359966. 0.0523	171768. 0.0459	599599. 0.0313	88418. 0.1844	264860. 0.0759
Bauru	120601. 0.0175	63416. 0.0170	274224. 0.0143	33316. 0.0695	89702. 0.0257
S. José R. Preto	159117. 0.0231	83997. 0.0225	104887. 0.0055	26515. 0.0553	143143. 0.0410
Araçatuba	82249. 0.0119	41833. 0.0112	78803. 0.0041	22206. 0.0463	84408. 0.0242
Presidente Prudente	91562. 0.0133	48075. 0.0129	96924. 0.0051	23195. 0.0484	71477. 0.0205
Marília	103133. 0.0150	57922. 0.0155	99958. 0.0052	33048. 0.0689	107390. 0.0308
Total do Estado	6887308.	3739515.	19183473.	479513.	3488633.

FONTE: CESP

(1) Incluem: Iluminação Pública, Poderes Públicos, Empresas de Serviços Públicos Inderdepartamental e Tração Elétrica.

TABELA XVI

MEDIDA DE AFINIDADE
CONSUMO DE ENERGIA ELETRICA EM SAO PAULO

ANO DE 1977

TIPO	RESIDENCIAL	COMERCIAL	INDUSTRIAL	RURAL	DEMAIS CASOS
RESIDENCIAL	1,00000	,99920	,98166	,70535	,98706
COMERCIAL	,99920	1,00000	,97843	,68069	,98131
INDUSTRIAL	,98166	,97843	1,00000	,70429	,96690
RURAL	,70535	,68069	,70429	1,00000	,80650
DEMAIS CASOS	,99706	,98131	,96690	,80650	1,00000

TABELA XVII - ESTADO DE SÃO PAULO - REGIÕES ADMINISTRATIVAS -
CONSUMO DE ENERGIA ELÉTRICA POR CLASSE - 1978 .

REGIÕES	RES.	COMERC.	INDUSTR.	RURAL	DEMAIS CLASSES (1)
Grande São Paulo	4739344. 0.6323	2637873. 0.6523	11442063. 0.5367	20994. 0.0372	2190278. 0.5485
Litoral	466610. 0.0623	251904. 0.0623	2213154. 0.1038	9047. 0.0160	205345. 0.0514
Vale do Paraiba	254651. 0.0310	118063. 0.0292	1248043. 0.0585	16186. 0.0287	138076. 0.0346
Sorocaba	250021. 0.0334	115130. 0.0285	1692177. 0.0794	55901. 0.0990	177390. 0.0444
Campinas	787136. 0.1050	389690. 0.0964	3351774. 0.1572	188293. 0.3335	370751. 0.0928
Ribeirão Preto	356803. 0.0476	190674. 0.0472	671097. 0.0315	102367. 0.1813	357849. 0.0896
Bauru	134454. 0.0179	75796. 0.0187	309735. 0.0145	40323. 0.0714	105053. 0.0263
S.José R. Preto	191445. 0.0212	97515. 0.0241	107741. 0.0051	36776. 0.0651	142277. 0.0356
Araçatuba	92720. 0.0124	48146. 0.0119	82063. 0.0038	26234. 0.0465	117913. 0.0295
Presidente Prudente	116657. 0.0156	53433. 0.0132	94742. 0.0044	28113. 0.0498	65427. 0.0164
Marília	115322. 0.0154	65732. 0.0163	107490. 0.0050	40284. 0.0714	123008. 0.0308
Total do Estado	7495169.	4043956.	21320079.	564518.	3993367.

FONTE: CESP

(1) Incluem: Iluminação Pública, Poderes Públicos, Empresas de Serviços Públicos Interdepartamental e Tração Pública.

TABELA XVIII

MEDIDA DE AFINIDADE
CONSUMO DE ENERGIA ELETRICA EM SAO PAULO

ANO DE 1978

TIPO	RESIDENCIAL	COMERCIAL	INDUSTRIAL	RURAL	DEMAIS CASOS
RESIDENCIAL	1.00000	.99958	.97741	.71415	.99026
COMERCIAL	.99958	1.00000	.97359	.70058	.98901
INDUSTRIAL	.97741	.97359	1.00000	.71380	.95852
RURAL	.71415	.70058	.71380	1.00000	.77680
DEMAIS CASOS	.99026	.98901	.95852	.77680	1.00000

3.2 - GRÁFICOS E INTERPRETAÇÃO

A partir das medidas calculadas, apresentadas nas matrizes da secção anterior, podemos ordenar os grupos de acordo com o consumo.

Em todos os anos, o consumo de energia elétrica é máximo para a classe Industrial; o que pode ser visto observando-se as proporções. Fixando esta e usando as medidas de Afinidades, as classes podem ser ordenadas como se segue: Industrial; Residencial ; Comercial; Demais Classes e Rural.

Esta ordenação é obtida da seguinte maneira: fixamos o grupo que apresenta maior proporção e em seguida consideramos as classes em ordem decrescente de Afinidade, em relação à fixada.

Por exemplo, para o ano de 1970, a Afinidade entre os grupos Industrial e Residencial é igual a 0.98968; a Afinidade entre Industrial e Comercial é de 0.98653, etc..., obtendo a ordem acima.

Esta mesma ordem podemos observar através do gráfico, em que plotamos as Afinidades das outras classes em relação à Industrial, Gráfico IV.

Dentre as classes consideradas, somente a Rural se mantém bastante afastada das demais, como podemos observar, pelos gráficos, numerados de I a V, apresentados a seguir.

GRÁFICO I

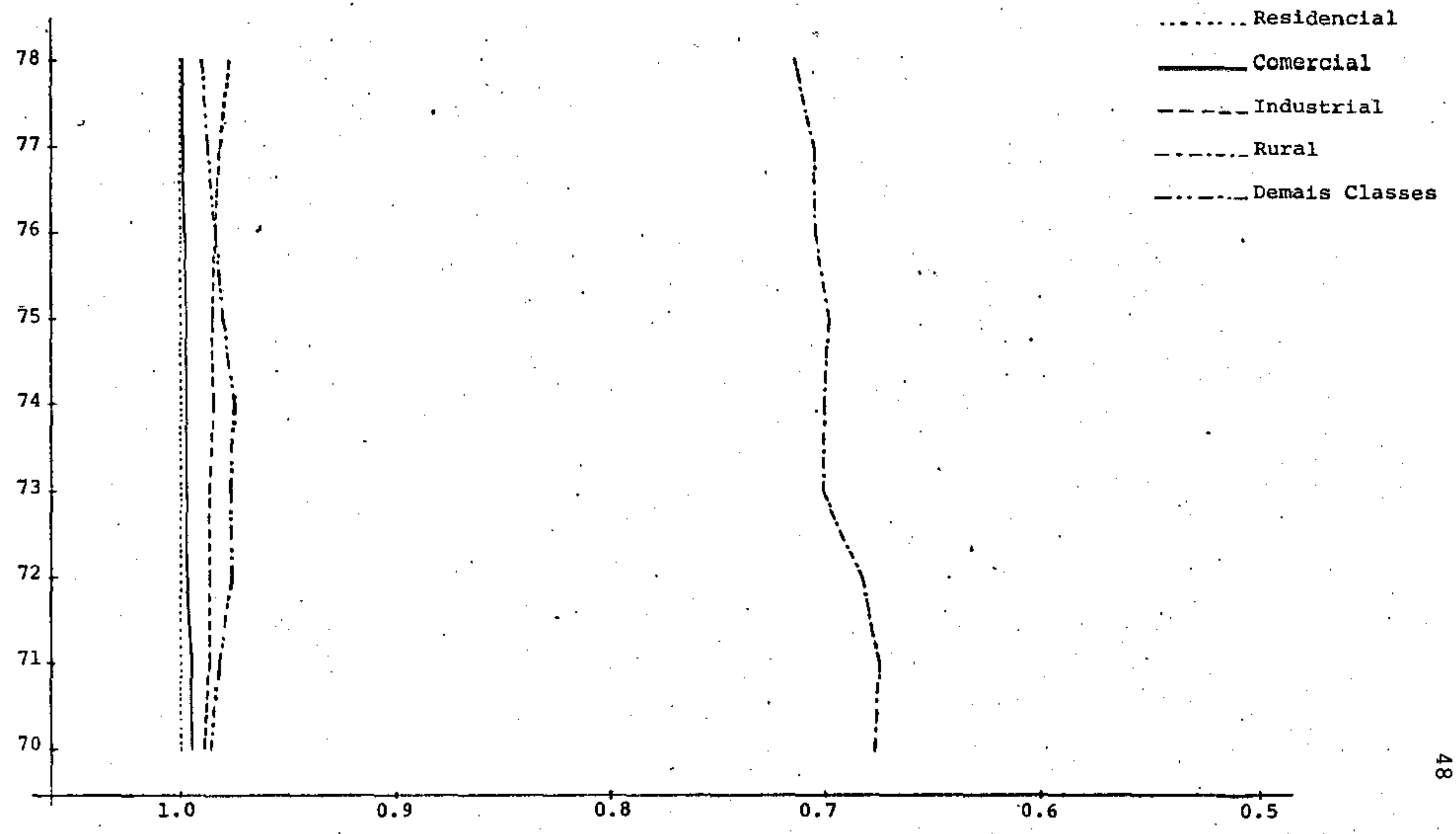

Residencial
Comercial
Industrial
Rural
Demais Classes
78
77
76
75
74
73
72
71
70
1.0
0.9
0.8
0.7
0.6
0.5

GRÁFICO II

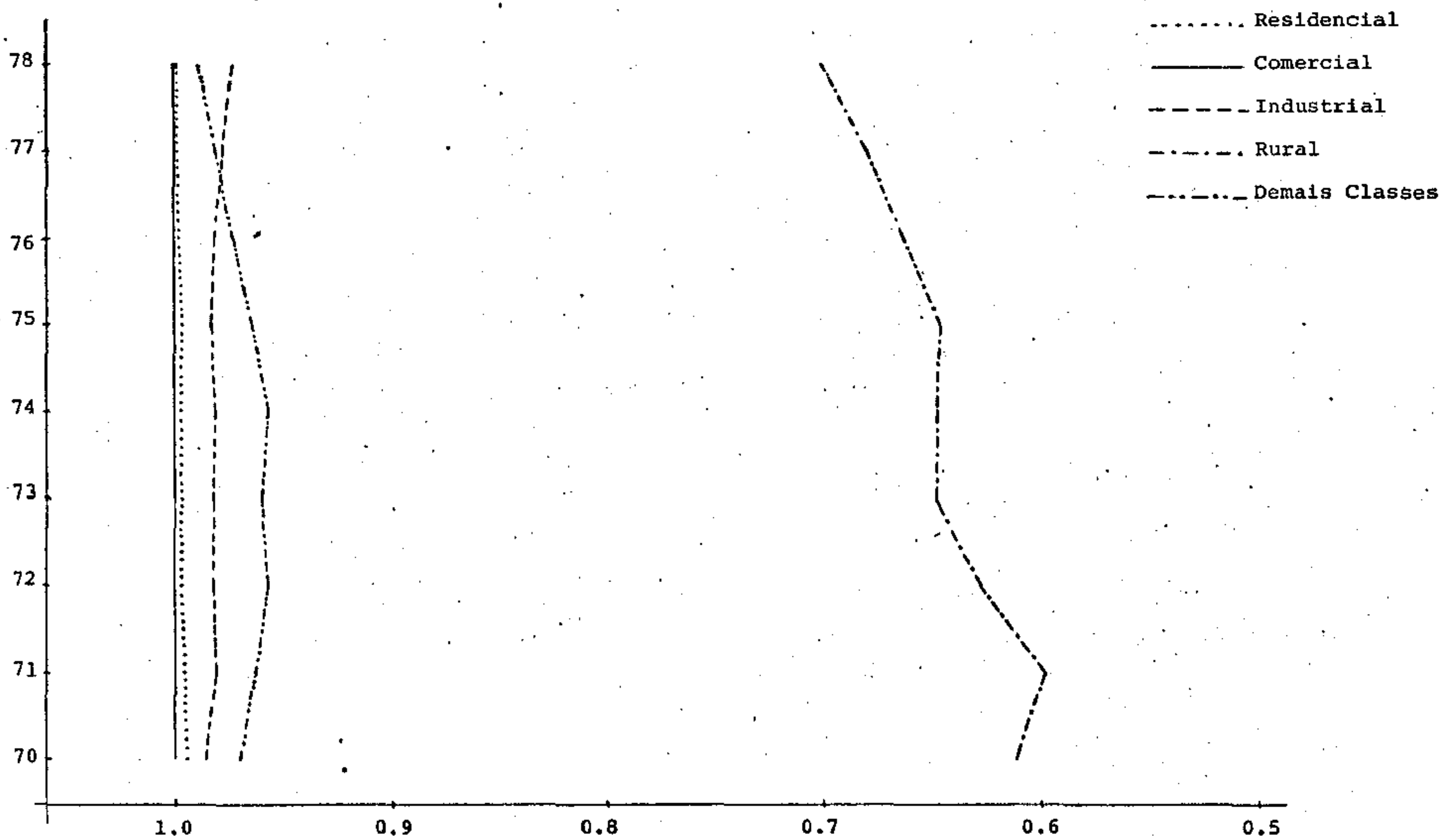

Residencial
Comercial
Industrial
Rural
Demais Classes
78
77
76
75
74
73
72
71
70
1.0
0.9
0.8
0.7
0.6
0.5

GRÁFICO III

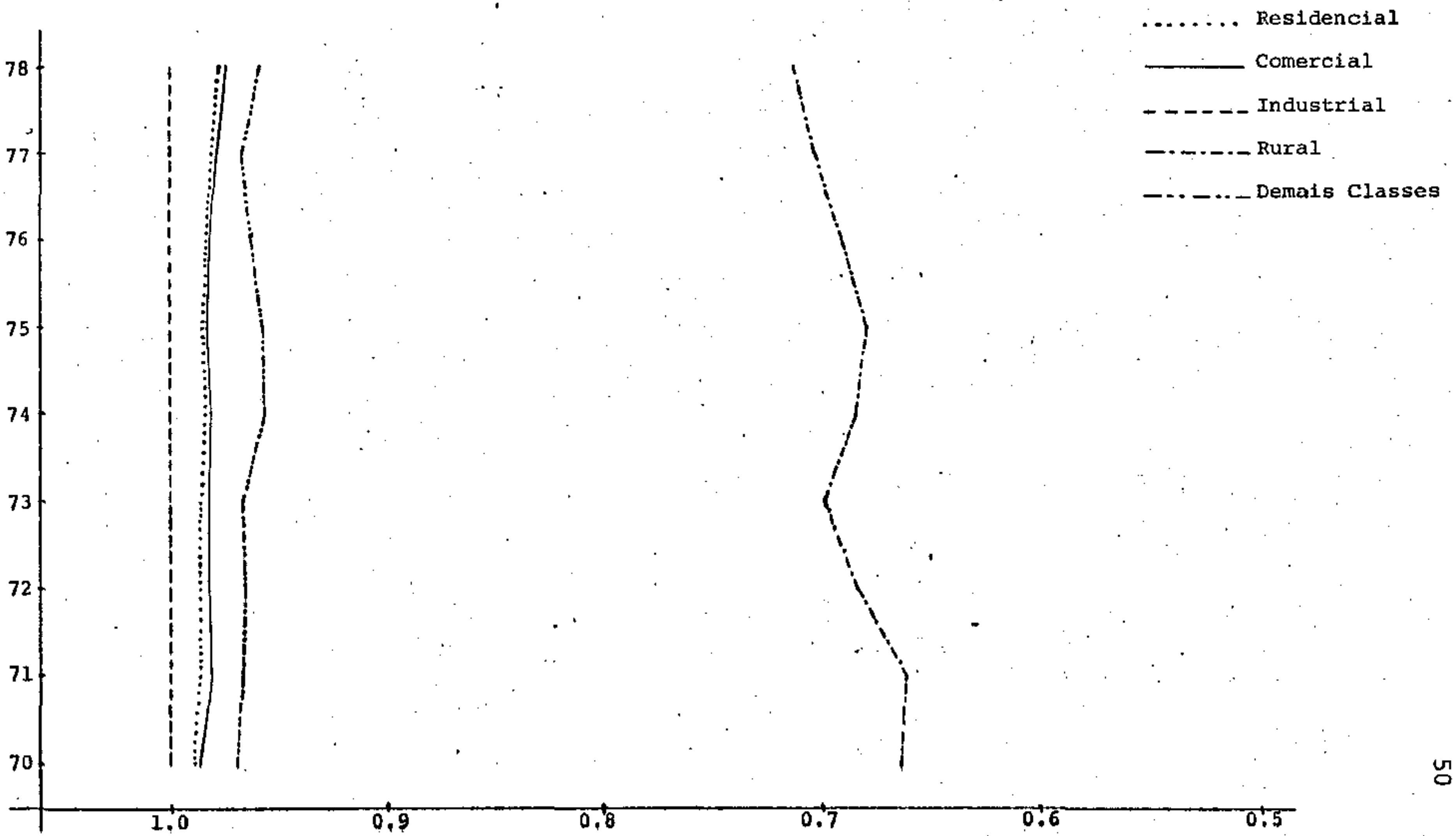

Residencial
Comercial
Industrial
Rural
Demais Classes
78
77
76
75
74
73
72
71
70
1,0
0,9
0,8
0,7
0,6
0,5

GRÁFICO IV

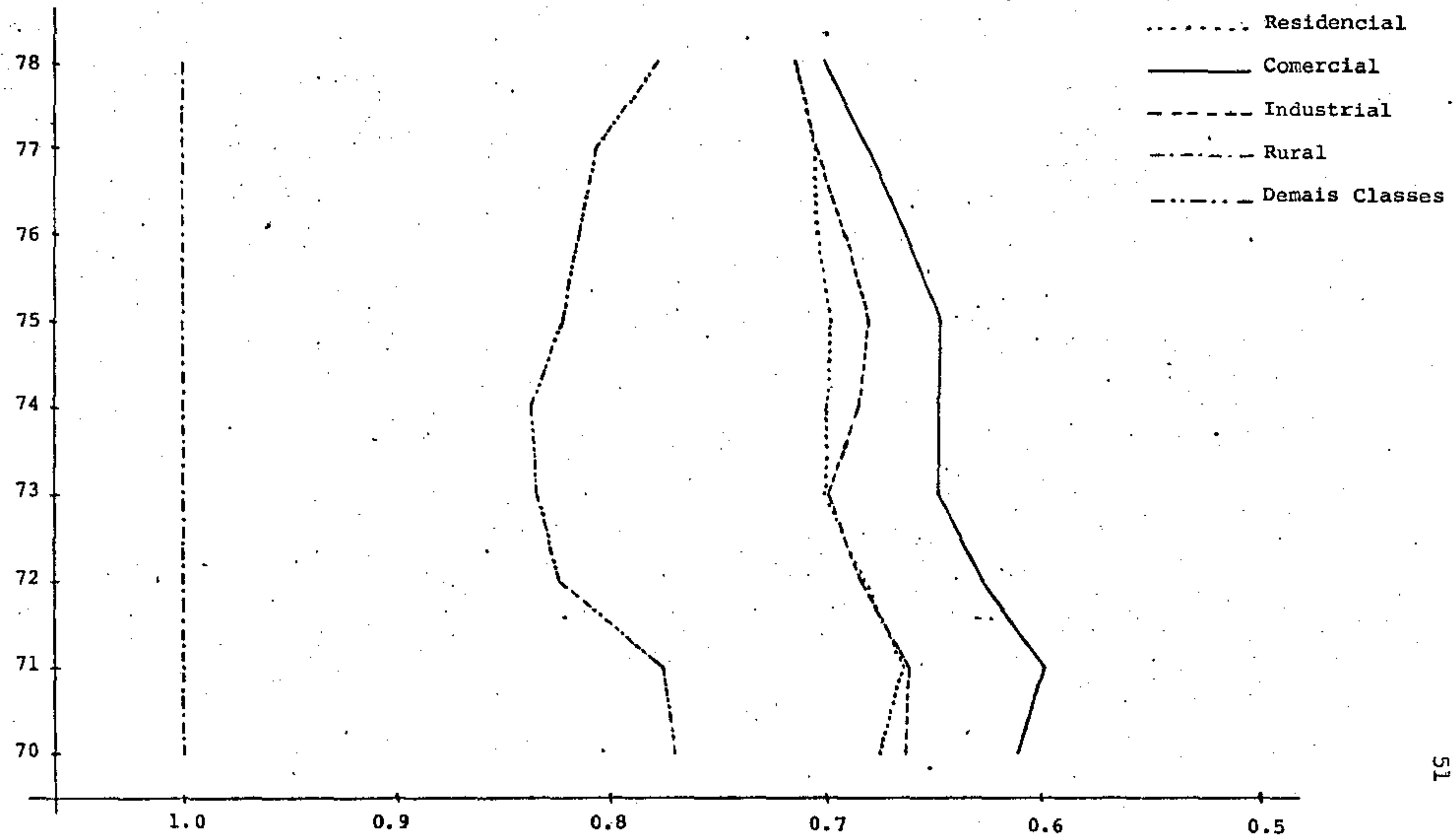
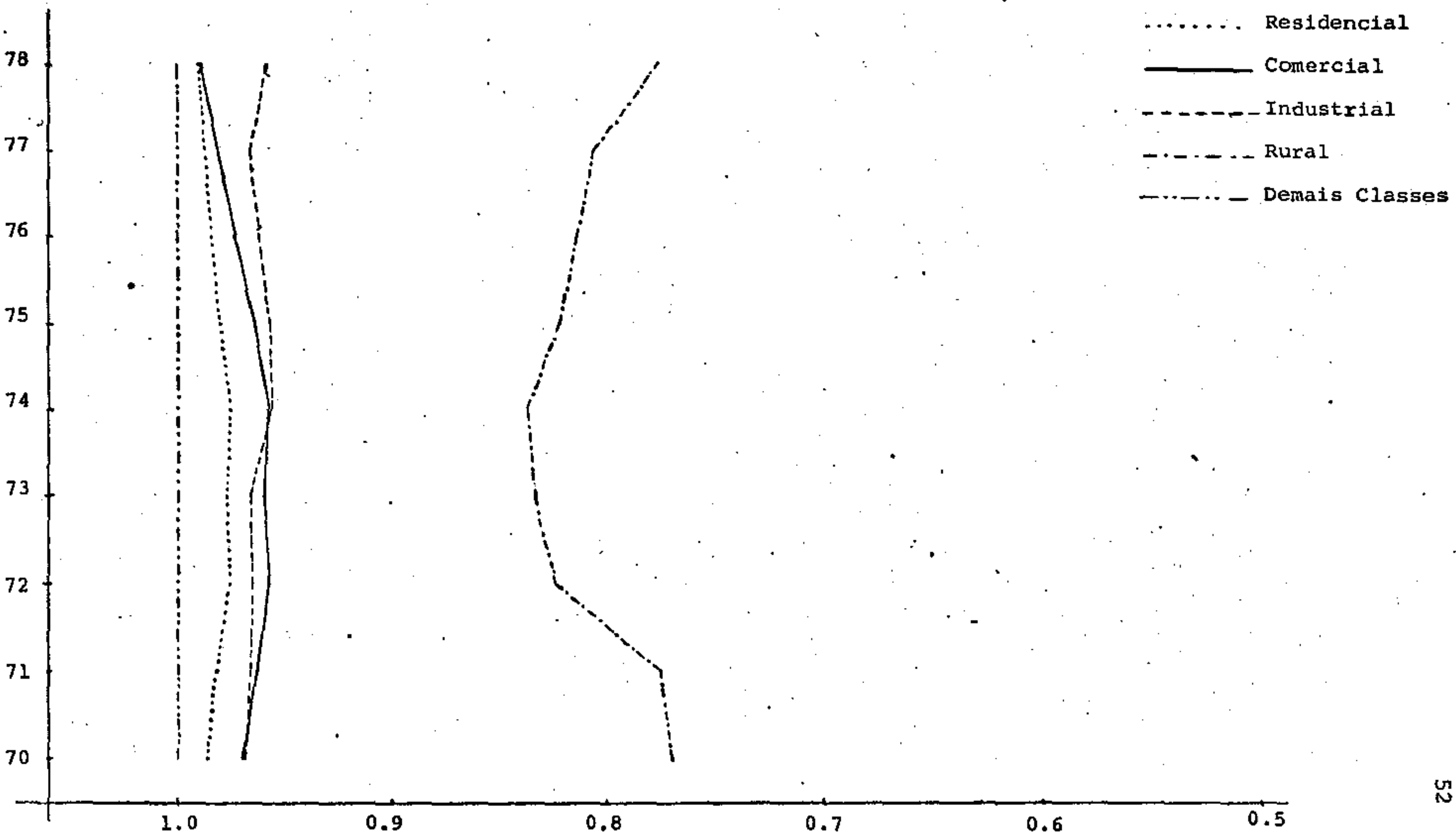

Residencial
Comercial
Industrial
Rural
Demais Classes
78
77
76
75
74
73
72
71
70
1.0
0.9
0.8
0.7
0.6
0.5

GRÁFICO V


Residencial
Comercial
Industrial
Rural
Demais Classes
78
77
76
75
74
73
72
71
70
1.0
0.9
0.8
0.7
0.6
0.5

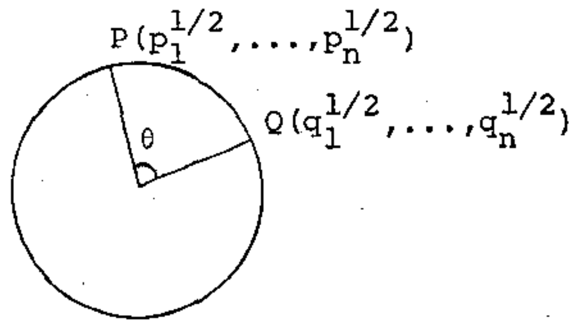
Como conclusão geral, podemos dizer que não há mudanças, neste período de 9 anos por nós estudado; ou seja, a ordem quanto ao consumo de energia elétrica por classe, nas Regiões Administrativas do Estado de São Paulo, se manteve.

CAPÍTULO IV

OUTROS RESULTADOS E APLICAÇÕES DA MEDIDADE DE AFINIDADE

A Afinidade mede a "proximidade" entre distribuições, ou seja, é uma medida que representa a discrepância entre distribuições, logo podemos usá-la como medida de discriminação. Assim sendo, pode ser largamente aplicável, principalmente em Inferência.

Geometricamente, a Afinidade pode ser interpretada como o cosseno do ângulo entre os pontos P e Q. Ou seja, se $P(p_1^{1/2},\ldots,p_n^{1/2})$ e $Q(q_1^{1/2},\ldots,q_n^{1/2})$ representam pontos únicos, em uma hiperesfera, então o cosseno do ângulo entre os dois vetores corresponde à Afinidade, logo $\cos\theta = \rho_n(P:Q)$.



Podemos ainda observar que, a distância entre os pontos P e Q é a Distância Matusita, que é definida por:

$$d_n(P:Q) = [\sum_{i=1}^{n} (p_i^{1/2} - q_i^{1/2})^2]^{1/2}$$

Esta medida pode ser escrita, também, como:

$$d_n(P:Q) = [2(1 - \sum_{i=1}^{n}(p_i q_i)^{1/2}]^{1/2}$$

$$d_n(P:Q) = [2(1 - \rho_n(P:Q))]^{1/2}$$

Ou sejam a Distância Matusita pode ser dada em função de A-finidade.

Bhattacharyya (1945-46) define uma medida de diverência, entre duas distribuições $(p_1,\ldots,p_n)$ e $(q_1,\ldots,q_n)$, como sendo o quadrado do ângulo entre os pontos $(p_1^{1/2},\ldots,p_n^{1/2})$ e $(q_1^{1/2},\ldots,q_n^{1/2})$; e, mostra que o cosseno deste ângulo é dado por

$\sum_{i=1}^{n}(p_i q_i)^{1/2}$.

Em vários desenvolvimentos teóricos ou aplicados, a Afinidade é usada juntamente com a distância de Matusita.

Matusita ((1954), (1955),(1967)) mostra como podemos usar a Distância, e consequentemente a Afinidade, em problemas de estimação de parâmetros, desconhecidos, de uma determinada distribuição.

Ainda sobre estimação de parâmetros, suficiência e outros problemas na área de Inferência, podemos encontrar outros resultados, que foram obtidos por Matusita ((1961),(1971)) e Kudo (1970).

Em problemas de classificação, Matusita (1957), estabeleceu regras de decisão, usando a Afinidade, para o caso em que são dadas duas variáveis aleatórias independentes, X , Y, e deseja-se saber se uma terceira variável aleatória Z, tem distribuição igual à de X ou igual à de Y, desde que uma delas seja válida. Os desenvolvimentos foram feitos para 2 casos: quando se tem várias observações e quando se tem apenas uma observação. E a regra de decisão estabelecida, pode ser extendida, facilmente, para várias variáveis. Ainda em ajuste de curvas, usando a Afinidade, podemos encontrar outros resultados obtidos por Matusita (1955).

Regras de decisão, para problemas de independência, podem ser encontradas em Matusita e Akaike (1956).

Alguns resultados, para a medida de Afinidade entre distribuições contínuas, foram apresentados por Matusita ((1961),(1967)).

Na área de Análise Multivariada, vários resultados foram obtidos por Matusita ((1957),(1966),(1967)).

Mathai e Rathie (1972) apresentam dois tipos de caracterização da Afinidade e da Distância, entre duas populações discretas. Mostram, ainda, a possibilidade de caracterizações análogas, para outras medidas, como a distência de Bhattacharyya, ver Adke (1958) e Bhattacharyya (1945-46); a invariância de Jeffreys, ver Adke (1958) e Jeffreys (1961); a discrepância de Pearson , ver Adke (1958), e a medida de dispersão generalizada de Mathai, ver Mathai (1967).

Uma caracterização feita por Mathai e Rathie (1972), se assemelha bastante a que apresentamos no Capítulo II, diferenciando, apenas, no postulado de recursividade.

Toussaint (1974b)apresenta uma série de resultados para a Afinidade entre várias distribuições.

Kaufman e Mathai (1973), também apresentam uma caracterização axiomática, para a Medida de Afinidade entre várias distribuições. O teorema apresentado nesta caracterização, se baseia em um conjunto de quatro axiomas: Recursividade, Simetria, Normalização e Continuidade. Este último axioma se faz necessário, devido a maneira como é apresentada a Recursividade.

Na área de aplicações, em dados numéricos, George e Mathai (1974) estudaram, através do número de nascidos vivos, em cada grupo religioso, a tendência do comportamento dos grupos, em várias cidades de Kerala, sudoeste da India. Este estudo foi feito usando tanto a Afinidade como a Distância.

A Afinidade, juntamente com outras medidas de Informação , foi utilizada na seleção das variáveis independentes, mais relevantes, para a construção de modelos de previsão da área plantada e de previsão da produção por hectare de soja, milho e trigo, na região de Ijuí, no Estado do Rio Grande do Sul, ver Rathie e outros (1979).

Lazo e Rathie (1979) utilizaram a Distância Matusita e a Afinidade, em dados de casamentos, na grande São Paulo, de 1931 a 1974, para medir o grau de integração, no Brasil, de quatro grupos de imigrantes: espanhóis, italianos, japoneses e portugueses.

Os trabalhos citados nos dão uma idéia das possíveis aplicações, em outras áreas, diferentes das já mencionadas, possibilitando análises e conclusões, em qualquer outro campo de pesquisa.

REFERÊNCIA

- ADKE, S.R. - (1958) - A Note on Distance Between Two Populations, *Sankhya*, 19, 195 - 200.

- ASH, R. - (1965) - Information Theory, *John Wiley & Sons, Inc., New York.*

- ATTNEAVE, F. - (1969) - Applications of Information Theory to Psychology, *Henry Holt and Company, New York.*

- BHATTACHARYYA, A. - (1945-46) - On a Measure of Divergence Between Two Multinational Population, *Sankhya*, 7, 401 - 406.

- COORDENADORIA DE ANÁLISE DE DADOS - (1978) - Boletim de Dados Conjunturais - vol. 1, nº 4, out/dez - *Governo do Estado de São Paulo - Secretaria de Economia e Planejamento.*

- FISHER, R.A. - (1948) - Statistical Methods for Research Workers, *10th Edition, Oliver-Boyd Ltd., Edimburg - London.*

- GEORGE, A. e MATHAI, A.M. - (1974) - Applications of the Concepts of Affinity and Distance to Population Problems, *J. Biosocial Sciences*, 6, 347 - 356.

- HARTLEY; R.V.L. - (1928) - Transmission of Information, *Bell System Tech. J.*, 7, 535 - 563.

- JEFFREYS, H. - (1961) - Theory of Probability, *3rd ed., Oxford at the Clarenden Press.*

- KAUFMAN, H. e MATHAI, A.M. - (1973) - An Axiomatic Foundation for a Multivariate of Affinity among a Number of Distributions, *Journal of Multivariate Analysis*, 3, 236 - 242.

- KUDO, H. - (1970) - On a Approximation to a Sufficient Statistics Including a Concept of Asymtotic Sufficiency, *Jour. Faculty of Science, Univ. of Tokyo*, Section 1, 17, 273-290.

- KULLBACK, S. - (1959) - Information Theory and Statistics, *John Wiley, New York*.

- LAZO, A.V. e RATHIE, P.N. - (1979) - Medidas de Informação: Algumas Aplicações Práticas, *Mimeo*. Apresentado na Xª Conferência Internacional de Biometria, 6-10 de Agosto, 1979, Guarujá, SP.

- MATHAI, A.M. - (1967) - Dispersion Theory, *Estadística*, 95, 271 - 284.

- MATHAI, A.M. e RATHIE, P.N. - (1972) - Characterization of Matusita's Measure of Affinity, *Ann. Inst. Statist. Math.*, 24, 473 - 483.

- MATHAI, A.M. e RATHIE, P.N. - (1975) - Basic Concepts in Information Theory and Statistics, Axiomatic Foundation and Applications, *Wiley, New York*.

- MATUSITA, K. - (1954) - On the Estimation by the Minimum Distance Method, *Ann. Inst. Statist. Math.*, 5, 59 - 65.

- MATUSITA, K. - (1955) - Decision Rules Based on the Distance, for Problem of Fit, Two Samples and Estimation, *Ann. Math. Statist.*, 26, 631 - 640.

- MATUSITA, K. - (1957) - Decision Rule , Based on the Distance, for the Classification Problem. *Ann. Inst. Statist. Math.*, 8, 67 - 77.

- MATUSITA, K. - (1961) - Interval Estimation Based on the Notion of Affinity, *Bull. International Statist. Inst.*, 38, 241 - 244.

- MATUSITA, K. - (1966) - A Distance and Related Statistics in Multivariate Analysis (Ed. by P.R: Krishnaiah), *Academic Press, New York*, 187 - 200.

- MATUSITA, K. - (1971) - Some Properties of Affinity and Applications, *Ann. Inst. Statist. Math.*, 23, 137 - 155.

- MATUSITA, K. e AKAIKE, M. - (1956) - Decision Rules, Based on the Distance, for the Problems of Independence, Invariance and Two Samples, *Ann. Inst. Statist.Math.*, 7, 67 - 80.

- RATHIE, P.N. e OUTROS - (1979) - Modelos para Previsão de Safra de Soja, Trigo e Milho na Região da COTRIJUI (R.G. S.) - COTEDEC - *Unicamp*.

- RÊNYI, A. - (1970) - Probability Theory, *North Holland Publishing Company, Amsterdam, London.*

- SHANNON, C.E. - (1948) - A Mathematical Theory of Communication, *Bell System Tech. J.*, 27, 379 - 423, 623 - 656.

- THEIL, M. - (1967) - Economics and Information Theory, *North Holland Publishing Company, Amsterdam e Randy McNally & Company, Chicago.*

- TOUSSAINT, G.T. - (1974b) - Some Properties of Matusita's Measure of Affinity of Several Distributions, *Ann. Inst. Statist. Math.*, (?), 389 - 394.

- WIENER, N. - (1948) - Cybernetics or Control and Communication in the Animal and the Machine, *Act. Sci. Indust.*, *Nr. 1053, Hermann et Cie, Paris.*