UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO
DEPARTAMENTO DE ENGENHARIA DE SISTEMA

ABORDAGEM BIPOLAR DO PROBLEMA DE CLASSIFICAÇÃO E ESCOLHA

Autor: Leandro Sauer
Orientador: Prof. Dr. Jurandir F. R. Fernandes
Co-Orientador: Prof. Dr. Sebastião de Amorim

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA - BAE - UNICAMP

Sauer, Leandro
Sa85a Abordagem bipolar do problema de classificação e escolha / Leandro Sauer.--Campinas, SP: [s.n.], 2003.

Orientador: Jurandir F. R. Fernandes e Sebastião de Amorim.
Tese (Doutorado) - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1.Classificação. I. Fernandes, Jurandir F. Ribeiro. II. Amorim, Sebastião de. III. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. IV. Título.

UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO
DEPARTAMENTO DE ENGENHARIA DE SISTEMA

ABORDAGEM BIPOLAR DO PROBLEMA DE CLASSIFICAÇÃO E ESCOLHA

Autor: Leandro Sauer
Orientador: Prof. Dr. Jurandir F. R. Fernandes
Co-Orientador: Prof. Dr. Sebastião de Amorim

Banca Examinadora:
Prof. Dr. Hermano Tavares
Prof. Dr. Luiz Koodi Hotta
Prof. Dr. Raul Vinhas
Prof. Dr. Reinaldo Castro Souza
Prof. Dr. Takaaki Ohishi
Prof. Dr. Akebo Yamakami

Tese de Doutorado apresentada à Faculdade de Engenharia Elétrica e de Computação (FEEC), da Universidade Estadual de Campinas (UNICAMP), como parte dos requisitos exigidos para obtenção do título de Doutor em Engenharia Elétrica**.**

Junho – 2003
Campinas – SP

AGRADECIMENTOS

Ao Prof. Dr. Sebastião de Amorim pela orientação segura, incentivo constante e amizade. Também por me ensinar que a solução dos problemas, na maioria das vezes, é simples e exige poucos recursos.

Ao Prof. Dr. Jurandir J. F. Fernandes pelos incentivos constantes e ter possibilitado institucionalmente este trabalho.

A banca examinadora pelas contribuições.

Aos amigos da TECNOMÉTRICA Estatística Ltda, Inês, Malu, Marlene, Rogério, Carlão e Ricardo pela amizade, solidariedade e apoio em todos os momentos de convívio.

Ao Prof. Dr.Amaury de Souza, Pró-Reitor de Pesquisa e Pós-Graduação da UFMS, pela sua compreensão e apoio no cumprimento dos prazos legais para a finalização deste trabalho.

Aos meus pais, Lucidio e Nilva, porque, como pais, seguem acreditando no filho e servindo como exemplo de amor, retidão e simplicidade, mostrando que as amizades conquistadas são o maior presente. Além disso, por proporcionar, apesar de todas intempéries, um ambiente produtivo para mim e para Inesila.

A minha sogra Maria Lídia Guedes Montenegro pelo constante incentivo e carinho.

A minha mulher, Inesila Montenegro Sauer, pelo carinho, amor, compreensão e habilidade em "tocar" a vida familiar durante as várias e muitas vezes longas ausências. Também por me mostrar que a auto-estima " remove montanhas".

Aos meus filhos, Leon e Luca, pela alegria de viver.

Aos amigos Jair Biscola e José Roberto Zorzatto pelo constante apoio, incentivo e amizade que em muitos momentos proporcionaram tranqüilidade para a realização deste trabalho.

Ao amigo Marcelo Martins dos Santos pela sua hospitalidade.

ABSTRACT

The problem of ranking and choice is widespread in a modern society. Often, the objects in a given set are ranked on a more or less formal way – from nations to candidates to a low ranking job – usually to orient the choice process of a group of subjects. Almost always the approach to ranking is one sided, in the sense that only characteristics of the objects being ranked are taken into consideration in the construction of the ranking function. Variables related to the subjects to whom the ranking is being performed is seldom involved. In this project we argue in favor of a two sided approach to the problem, where variables related to both objects and subjects are pooled together into a bipolar statistical model, in which the score of an object depends not only of its properties, but also, and on an equal status, of characteristics of the subject by whom – and to whom – the ranking is being done. In this project we develop the theoretical framework for the construction of bipolar ranking models. Two different experimental designs for data collecting are presented and fully analyzed. One allows for the use of the powerful statistical methods for multiple linear regression for estimating the parameters; the other, more generally applicable, had to be treated by the maximum likelihood criterion, and therefore demanded intense computational involvement, and the application of the asymptotic results of the Fisher Information Theory for inference on the estimators. Several other experimental approaches were suggested throughout the text. Bipolar model estimation is often a computationally intensive job. In this project we developed the basic structure of the problem of estimating parameters by maximizing the associated log-likelihood function, and applied it in some numeric examples. The basic developments suggest several interesting lines for further development, both in terms of design of experiment and of efficient numerical strategies for parameters estimation.

RESUMO

O problema de classificação e escolha é comum na sociedade moderna. Freqüentemente, objetos são classificados de uma maneira mais ou menos formal – desde nações a candidatos a um emprego – usualmente para orientar o processo de escolha de um grupo de sujeitos. Quase sempre a abordagem é unipolar, no sentido que somente características dos objetos classificados são consideradas na construção da função de classificação. Variáveis relacionadas aos sujeitos para quem a classificação é construída raramente são envolvidas. Neste trabalho, nós argumentamos em favor de uma abordagem bipolar para este problema, onde variáveis relacionadas aos sujeitos e objetos, conjuntamente, são consideradas dentro de um modelo estatístico bipolar, no qual o escore de um objeto depende não somente de suas propriedades, mas também, em igual importância das características dos sujeitos para o quais a classificação esta sendo feita. Neste projeto, nós desenvolvemos a estrutura teórica para a construção de modelos de classificação bipolar. Dois diferentes modelos de coleta de dados são apresentados e completamente analisados. Um permite o uso de todo instrumental estatístico utilizado na estimação de parâmetros em modelos de regressão linear múltipla; o outro, mais geralmente aplicável, tem que ser tratado por critérios de máxima verossimilhança, e portanto demanda intenso esforço computacional e a aplicação de resultados assintóticos da teoria de informação de Fisher para inferência de seus estimadores. Várias outras abordagens experimentais são sugeridas através do texto. A estimação de modelos bipolares emprega freqüentemente intensivo esforço computacional. Neste trabalho nós desenvolvemos a estrutura básica de estimação dos parâmetros pela maximização da função de log-verossimilhança e aplicamos isto em alguns exemplos numéricos.

INTRODUÇÃO GERAL

INTRODUÇÃO AO PROBLEMA DE CLASSIFICAÇÃO E ESCOLHA

A cada dois anos a MacMillan[1] publica sua classificação das maiores áreas metropolitanas americanas segundo o critério do Índice de Qualidade de Vida Urbana (IQVU). Na sua edição de 1999[2], divulgada em novembro daquele ano, as 351 principais áreas metropolitanas dos Estados Unidos foram classificadas; a área metropolitana de Orange County, na Califórnia, ficou em primeiro lugar, com Seattle em segundo. Na Tabela 1 apresentamos as dez melhores e as dez piores segundo aquela classificação.

Tabela 1 - As dez melhores e as dez piores áreas metropolitanas dos Estados Unidos, segundo classificação pelo IQVU, do Places Rated Almanac, edição de 1999.

Área Metropolitana	Classificação
Orange County, CA	1
Seattle - Bellevue - Everett, WA	2
Houston, TX	3
Washington, DC - MD - VA - WV	4
Phoenix - Mesa, AZ	5
Minneapolis - ST.Paul , MN - WI	6
Atlanta, GA	7
Tampa - St. Petersburg - Clearwater, FL	8
San Diego, CA	9
Philadelphia, PA - NJ	10

Área Metropolitana	Classificação
Yuma, AZ	342
Albany, GA	343
Sumter, SC	344
Lawton, OK	345
Lima, OH	346
Jackson, MI	347
Vineland - Millville - Bridgeton, NJ	348
Dover, DE	349
Elmira, NY	350
Mansfield, OH	351

Este esforço de classificação vem sendo empreendido sistematicamente, por aquela instituição, desde 1985, e é importante sobre vários aspectos. Por um lado, serve de orientação básica para pessoas interessadas em mudar de cidade – uma operação muito freqüente numa sociedade com a mobilidade geográfica da americana – e para as grandes corporações, na localização de seus escritórios regionais e mesmo de suas sedes. Por outro lado, tanto as administrações municipais, quanto órgãos dos governos estadual e federal, podem contar com um termômetro objetivo de boa qualidade para a aferição, no nível local, da eficácia de políticas e de programas aplicados. (Ver mais sobre este tópico no Capítulo II e Apêndice A: apresentação do questionário e Dados Brutos do IQVU).

[1] Diversas outras companhias e órgãos governamentais dos EUA publicam listas semelhantes.
[2] SAVAGEAU, David and LOFTUS, GEOFFREY, Loftus **Places Rated Almanac 5th Edition** Macmillan New York , USA 1997

A abordagem adotada para a ordenação de áreas urbanas pelo critério do índice da qualidade de vida é interessante, mas demanda alguma reflexão. Primeiro vale notar que qualquer estratégia eficaz de classificação de áreas urbanas será sempre, inevitavelmente, multicritérios, uma vez que o índice de qualidade de vida é uma combinação mais ou menos complexa de uma variedade de quesitos, tais como segurança pessoal e patrimonial, sistemas de educação e de saúde, facilidade de transporte, oportunidades de emprego, características do clima, custo de vida, opções de lazer, entre outros.

Abrem-se aqui algumas questões importantes. Como medir, por exemplo, para uma dada área urbana, o Índice de Segurança do Cidadão? Uma vez estabelecidas definições adequadas para cada indicador, o mesmo deverá ser medido. Com um vetor de indicadores já quantitativamente definido para cada área urbana considerada, como consolidar estas componentes num único escalar – o Índice de Qualidade de Vida – para aquela área urbana? Em outras palavras, qual a relação de barganha entre, por exemplo, Segurança e Oportunidade de Emprego? Isto é, quanto de segurança o cidadão está disposto a trocar por cada ponto extra do nível de Oportunidade de Emprego?

Na classificação de candidatos à uma determinada faculdade, quantos pontos na nota de Física se barganha por cada ponto em Biologia? Classificação pela média aritmética assume implicitamente uma barganha um por um, radicalmente arbitrária. Médias ponderadas flexibilizam esta rigidez, mas os pesos, são geralmente escolhidos de forma arbitrária, com direções consensuais (em Medicina, Física é menos importante que Biologia) mas, a relação quantitativa de barganha é raramente investigada com maior profundidade [3].

Questões desta natureza são centrais a toda estratégia eficaz de classificação multicritérios.

Uma vez definida a expressão do Índice de Qualidade de Vida em função do valor dos diversos índices adotados, o problema estará resolvido.

Esta abordagem não é, naturalmente, inédita, e tem sido aplicada neste e em diversos outros contextos, em todo o mundo. No Brasil, por exemplo, alguns periódicos e a OAB, entre outras organizações, têm publicado *rankings* das melhores faculdades brasileiras[4] ou das melhores empresas para se trabalhar[5], entre outros tópicos.

Inúmeros artigos nos mais diversos jornais científicos do mundo, tais como: International Journal of Social Economics [34], American Demographics Magazine [28], The American Statistician [8], Social Indicators Research [45], Journal of the Royal Statistical Society [33], entre outros, mostram que profissionais das mais diversas áreas (economistas, estatísticos, administradores de empresa, sociólogos, psicólogos, antropólogos, entre outros) têm abordado este problema. Contudo, as discussões resumem-se aos critérios avaliados para cada país, estado, área metropolitana, cidade, faculdade ou empresa, sem nenhuma discussão sobre a definição da estrutura da função escore utilizada para a classificação.

Podemos imaginar como foi a discussão para a definição dos pesos das diferentes disciplinas, para os vestibulares à diversas faculdades da UNICAMP, é certo que a base tenha sido sempre qualitativa, e com pesos exatos definidos de forma essencialmente arbitrária.

Classificação de organizações, escolas, cidades ou países, mais que refletem, criam reputações. Num interessante artigo, Baden-Fuller, Ravazzolo e Schweizer [1], comentam este tipo de fenômeno, que vem ocorrendo na classificação de escolas de gestão de negócio (Business Schools) européias. Por este motivo, os processos de classificação devem receber um tratamento cuidadoso, em todos os seus aspectos: definição dos critérios, maneira de medi-los e como processá-los.

---

[3] No vestibular a UNICAMP, a partir de 1995, para Medicina, Biologia tem peso duas vezes que Matemática, a relação se inverte, por exemplo, para as Engenharias.
[4] Revista Playboy edição nº 302 , mês de Setembro de 2000.
[5] Revista Exame, edição nº 721 23/08/2000. (Reportagem de Capa)

Desde 1994, a ONU tem publicado anualmente uma ordenação dos países do mundo segundo o Índice de Desenvolvimento Humano (IDH)[6]. Seus critérios são baseados em três dimensões básicas do desenvolvimento humano: uma vida longa e saudável, acesso à informação e um padrão de vida decente. Estas três dimensões são combinadas com o mesmo peso numa função escore que induz uma ordenação dos países. No Relatório de 2002, o Brasil ocupa uma não muito honrosa 73ª posição, com a Noruega em primeiro e Serra Leoa em último, entre os 173 países avaliados. Para mais detalhes sobre como é calculado o Índice de Desenvolvimento Humano, ver **Apêndice B**.

Neste trabalho nós desenvolvemos um tratamento estatístico formal e rigoroso para o problema geral da classificação e escolha – do qual a classificação de áreas urbanas e países é um caso particular - a partir da constatação de limitações estruturais na abordagem tradicional para alguns casos deste tipo de problema.

O PROBLEMA GERAL DA CLASSIFICAÇÃO E ESCOLHA

De início vamos deixar um ponto bem claro: A natureza implicitamente bipolar do problema básico de classificação e escolha. Qualquer problema de classificação envolve dois grupos distintos de entidades: os Sujeitos e os Objetos.

No exemplo da classificação das áreas metropolitanas de um país, segundo o critério do índice de qualidade de vida, por exemplo, as áreas metropolitanas consideradas são os Objetos; os cidadãos, que vão se orientar pela classificação na formulação de decisões de escolha, são os sujeitos do problema. Nestes contextos os sujeitos são importantes, pois, para começo de conversa, é para eles que a classificação está sendo feita.

Como usualmente feitas e divulgadas na grande imprensa, a bipolaridade natural e inevitável do problema parece esquecida. Foca-se nos objetos, aparentemente esquecendo-se dos sujeitos, o que confere aos resultados uma natureza mais ou menos abstrata. Se

[6] Ver www.undp.org.br/HDR/HDR2002

Orange County, na Califórnia, foi considerada a melhor área metropolitana dos EUA na classificação de 1999, cabe aqui a pergunta: segundo os valores e preferências de quem?

Esta não é uma pergunta sem propósito. Pessoas têm gostos, preferências e prioridades diferentes, ainda que concordem em classificar cidades segundo o mesmo conjunto de critérios. Os pesos relativos de cada critério, contudo, podem e devem variar de pessoa a pessoa ou, numa métrica mais grossa, de grupo social a grupo social.

Se o critério "Ofertas de Emprego" é muito importante para jovens profissionais em início de carreira, ele pode significar pouco para cidadãos de meia idade, recém aposentados, que provavelmente estarão mais interessados em centros urbanos com variada oferta de opções lazer, alto nível de segurança e abundante oferta de serviços de transporte e saúde de boa qualidade.

A pergunta: **"**Aos valores e preferências de qual grupo social (etário, econômico, cultural) está direcionada a função escore adotada?", deve ser discutida cuidadosamente. Se o foco fosse em um certo "perfil médio da população" – americana, no caso – ficaria então no ar a sensação de que, talvez, aquela classificação apresentada na imprensa não diga respeito especificamente a ninguém ou, na melhor das hipóteses, a um subconjunto muito específico daquela população.

Neste trabalho nós construímos uma abordagem – conceitual, estatística e computacional completa - para o problema da Classificação e Escolha, que parte da constatação de que o mesmo é, estruturalmente, bipolar: a função escore deverá envolver não apenas variáveis associadas aos objetos, mas também variáveis associadas aos sujeitos. Em outras palavras, e ainda no exemplo citado acima, da classificação das áreas urbanas americanas quanto ao critério do Índice de Qualidade de Vida, o escore que um dado indivíduo associará a uma dada área urbana dependerá de variáveis da área urbana, tais como Segurança, Sistemas de Educação e de Saúde, Facilidade de Transporte, Oportunidades de Emprego, Características do Clima, Custo de Moradia, Opções de Lazer,

e também de variáveis do indivíduo, como Idade, Sexo, Estado Civil, Nível de Instrução, Profissão, entre outras.

A partir desta base estrutural, construiremos modelos de classificação de objetos para sujeitos. A posição relativa de um objeto no conjunto de objetos será não mais uma constante, abstratamente independente do particular indivíduo a que se orienta, mas variará em função do indivíduo ao qual se refere à classificação.

O MODELO BÁSICO

Nosso modelo envolve duas entidades básicas principais: o conjunto dos sujeitos, $\boldsymbol{S}$, e o conjunto dos objetos, $\boldsymbol{O}$. Os elementos de $\boldsymbol{S}$ e de $\boldsymbol{O}$ são caracterizados pelos vetores $\underset{\sim}{W}$ e $\underset{\sim}{X}$, respectivamente. A abordagem usual do problema, assumindo implicitamente que $\boldsymbol{S}$ é um conjunto unitário, é um caso particular muito específico que, contudo, pode ser adequado a diversas situações de interesse.

Admitimos uma Função Escore, $\mathscr{S}$, que associa um escalar positivo a cada par $(s , o) \in (\boldsymbol{S} \times \boldsymbol{O})$, dada por

$$\mathscr{S} = \mathscr{S}\left[\underset{\sim}{w}(s) , \underset{\sim}{x}(o) \right]$$

O escore real que um dado sujeito associa a um dado objeto será igual a $\mathscr{S}$, mais um desvio δ, decorrente de particularidades tanto do sujeito quanto do objeto, que não estão cobertas pelos vetores de características $\underset{\sim}{W}$ e $\underset{\sim}{X}$. Assim, o escore que um dado sujeito $\boldsymbol{s} \in \boldsymbol{S}$ associa a um dado objeto $\boldsymbol{o} \in \boldsymbol{O}$ será representado por uma variável aleatória $\mathscr{Y}$, definida como

$$\mathscr{Y} = \mathscr{Y}(\boldsymbol{s}, \boldsymbol{o}) = \mathscr{S}\left[\underset{\sim}{w}(s) , \underset{\sim}{x}(o) \right] + \delta(\boldsymbol{s}, \boldsymbol{o})$$

onde $E[\delta/(s,o)]$ tem esperança 0 e variância finita. Na expressão acima, $\mathscr{S}\left[\underset{\sim}{w}(s) , \underset{\sim}{x}(o) \right]$ corresponde à componente genérica do escore, comum a todos os pares $(\boldsymbol{s}, \boldsymbol{o}) \in \boldsymbol{S} \times \boldsymbol{O}$ que

partilham dos mesmos vetores $\underset{\sim}{w}(s)$ e $\underset{\sim}{x}(o)$. A componente $\delta(\boldsymbol{s}, \boldsymbol{o})$ constitui a componente do escore que é específica ao par $(\boldsymbol{s}, \boldsymbol{o})$.

Neste trabalho desenvolveremos diversas estratégias para se estimar a componente genérica, $\mathscr{S}\left[\underset{\sim}{w}(s), \underset{\sim}{x}(o)\right]$, a partir de dados experimentais.

MODELAÇÃO DA FUNÇÃO ESCORE $\mathscr{S}$

A função escore será modelada por funções polinomiais de grau nunca superior a dois. Em particular, trataremos exclusivamente da família de modelos em que Y é uma função polinomial completa, de ordem um ou dois, de $x_1, x_2, \ldots, x_v$, cujos parâmetros são funções lineares de $w_1, w_2, \ldots, w_u$.

Estas restrições, usuais na construção de modelos de superfícies de respostas, não comprometem significativamente nossa flexibilidade de modelação se superfícies suaves, como as envolvidas em problemas da natureza abordada.

Neste contexto, com m = n = 2, isto é, uma função de grau dois e duas variáveis, o modelo desta família será:

$$\mathscr{S}\left[\underset{\sim}{w}(s), \underset{\sim}{x}(o)\right] = a_o + a_1x_1 + a_2x_2 + a_{12}x_1x_2 + a_{11}x_1^2 + a_{22}x_2^2$$

com $a_i = b_{i,0} + b_{i,1}w_1 + b_{i,2}w_2$, para todo i. Este modelo envolve um total de 18 parâmetros.

CONDIÇÕES EXPERIMENTAIS

O modelo será estimado a partir de dados experimentais. Neste trabalho trataremos de algumas situações alternativas com relação à natureza e condições gerais de obtenção destes dados. Primeiramente, vamos considerar situações em que, confrontado com um

objeto $\boldsymbol{o} \in \boldsymbol{O}$, qualquer sujeito $\boldsymbol{s} \in \boldsymbol{S}$ será capaz de fornecer uma avaliação objetiva, embora imperfeita, do escore absoluto $\mathscr{Y}(\boldsymbol{s}, \boldsymbol{o})$, representado por $y(\boldsymbol{s}, \boldsymbol{o})$, com

$$y(\boldsymbol{s}, \boldsymbol{o}) = \mathscr{Y}(\boldsymbol{s}, \boldsymbol{o}) + \varepsilon$$

onde ε é um erro aleatório, com esperança 0 e variância finita, resultante da agregação do desvio $\delta(\boldsymbol{s}, \boldsymbol{o})$ com o erro de avaliação. Como não é nosso objetivo, neste trabalho, tratarmos dos desvios por especificidade, mas apenas do cerne genérico da função escore, a partir deste ponto não mais nos referiremos a δ isoladamente, mas apenas na forma agregada ao erro de avaliação, ε.

Nestes casos, técnicas clássicas de delineamento de experimento, bem como de ajuste do modelo, por regressão linear, serão utilizadas.

Noutro contexto, o sujeito não consegue avaliar o escore absoluto de um objeto, mas apenas ordenar, com imperfeições, naturalmente, dois objetos apresentados. Para estes casos desenvolveremos uma modelação para a estrutura de erro da ordenação, e uma estratégia de ajuste de modelos pelo critério da razão de verossimilhança.

Como os modelos geralmente envolvem um número elevado de parâmetros, os problemas numéricos de otimização oferecem desafio computacional considerável. Esta barreira inibiu, no passado, a exploração de modelos destes níveis de complexidade, em virtude da escassez de recursos computacionais. O extraordinário desenvolvimento dos computadores nas últimas décadas – que coloca hoje, na escrivaninha do pesquisador, máquinas que há vinte anos atrás seriam classificadas como super-computadores – subverteu dramaticamente a situação, embora, distraidamente, muitos pesquisadores ainda não tenham acordado para esta nova realidade.

Estratégias eficientes de busca serão empregadas, explorando características gerais de suavidade das superfícies de razão de verossimilhança. Resultados assintóticos relativos aos estimadores de máxima verossimilhança, referentes à distribuição assintótica e à matriz

de informação de Fisher, serão empregados para a determinação das propriedades estatísticas dos estimadores.

Em todos estes casos, apresentamos resultados teóricos para a qualidade do ajuste do modelo, bem como da ordenação produzida. Os resultados teóricos serão ilustrados através de simulações em computador e de algumas situações reais, através da comparação das propriedades observadas, com as previstas pelos resultados assintóticos.

Nos casos em que técnicas de regressão linear podem ser utilizadas, adotaremos sempre a suposição de normalidade dos erros para se construir inferências sobre os parâmetros. Contudo, a abordagem pela técnica *bootstrap* (Efron,B.e Tibshirani, R.J. [21]) pode também ser explorada nestes casos, como alternativa para quando a suposição de normalidade não puder ser adequadamente fundamentada.

CAPÍTULO I- O ÍNDICE DE QUALIDADE DE VIDA URBANA E O VESTIBULAR

Neste capítulo iremos mostrar, através de dois exemplos numéricos, todas as etapas envolvidas na classificação de objetos e as implicações da abordagem vigente. O primeiro exemplo refere-se à classificação de áreas metropolitanas nos EUA, segundo o Índice de Qualidade de Vida Urbana (IQVU), proposto por Savageau & Loftus [42] e o segundo a análise do concurso vestibular, exemplificando através dos dados do Vestibular da UNICAMP do ano de 1999, para os cursos de Engenharia Elétrica, Ciências Econômicas e Medicina.

O ÍNDICE DE QUALIDADE DE VIDA URBANA

Para classificarmos áreas metropolitanas segundo o IQVU, em primeiro lugar devemos definir quais os critérios que serão considerados. Savageau & Loftus consideram nove fatores:

- Custo da Moradia,
- Facilidade de Transporte;
- Oportunidades de Emprego;
- Sistema de Educação;
- Sistema de Saúde;
- Características do Clima;
- Segurança;
- Opções de Lazer e
- Opções Culturais.

Tais fatores cobrem substancialmente os vários aspectos da vida cotidiana e fundamentam as escolhas por um lugar para morar. Como comentamos na Introdução, as maiores discussões no ambiente da classificação, encontram-se neste ponto, isto é, na discussão dos critérios a serem considerados e como medi-los.

Neste contexto, no Brasil, podemos citar, como exemplo, duas referências históricas: o trabalho de LEMOS; ESTEVES e SIMÕES [31] e [43] de 1995, discutindo critérios e uma metodologia para medição da qualidade de vida urbana em Belo Horizonte, e o trabalho de BARBOSA [2] discutindo a própria postura de se definir critérios para medir a qualidade de vida urbana, de 1996.

No âmbito internacional a literatura é vasta,[7] porém restrita à discussão de critérios a serem utilizados. À função escore é dado geralmente um tratamento secundário, utilizando-se, em geral, a média aritmética dos fatores, como se pode ver em [3], [16]-[19], [26], [28], [32] - [34], [38], [39], [44].

Estes fatores precisam ser considerados segundo indicadores específicos, isto é, a necessidade de definir, por exemplo, o Custo da Moradia. Este é o segundo ponto a ser considerado no processo de classificação, ou seja, como medir cada um dos fatores. Em "*Places Rated Almanac"* de Savageau & Loftus, encontra-se a descrição pormenorizada dos indicadores para cada um dos nove fatores definidos acima.

Definidos os critérios, seus indicadores e a forma de medi-los, necessitamos definir como combiná-los. Como agregar num único indicador escalar os valores encontrados em cada indicador, muitas vezes com unidades e magnitudes diferentes? Por exemplo, no fator Facilidade de Transporte, teremos que combinar o suprimento de transporte público com a facilidade de conexão inter modal com outras áreas, envolvendo diversas companhias independentes e órgãos governamentais, através da rede viária. A solução adotada é a padronização de cada um dos indicadores, permitindo operações quantitativas entre os mesmos.

Desta forma temos, para cada fator, uma operação linear com seus indicadores padronizados, dependendo do peso de cada indicador, obtendo o escore geral correspondente. A média dos escores de cada um dos fatores utilizados para compor a qualidade de vida urbana das áreas metropolitanas, é utilizada para construir a classificação.

[7] Ver Bibliografia Organizada sobre Qualidade de Vida no site www.cob.vt.edu/market/isqols/bibres.htm

É importante lembrar que, ao usarmos a média dos escores para compor o escore geral, estamos implicitamente dando o mesmo peso, para cada um dos critérios considerados.

No anexo A, apresentamos os escores para cada fator por área metropolitana e a correspondente classificação obtida através do escore geral (média dos escores obtidos). Observe que este procedimento não considerou, em nenhum momento, os sujeitos para os quais a classificação está sendo feita. Neste procedimento está implícito que os objetos, e apenas eles, reúnem todas as informações necessárias à construção de um modelo de classificação de objetos para todos os sujeitos.

A classificação destas áreas metropolitanas considerando, especificamente, um sujeito, é abordada superficialmente através de um inventário sobre preferências, no qual são apresentadas as 72 situações envolvendo os 9 fatores tomados dois a dois, com o objetivo de verificar as prioridades de um sujeito especificamente. Assim ao final do inventário, temos para cada fator a quantidade de vezes que ele foi considerado prioritário, esta proporção é sugerida como o peso daquele fator para este sujeito.

Embora representando um avanço na direção bipolar, por considerar a possibilidade de pesos diferentes para sujeitos diferentes, esta abordagem é simplória e imperfeita. Numa ordenação perfeitamente consistente dos pares – o que não necessariamente ocorre, mesmo se tratando de uma ordenação cuidadosa – os pesos seriam iguais a 0, 1, 2, …, 8, respectivamente, para os 9 fatores ordenados por no sentido crescente do nível de prioridade para o usuário. Dizer que um fator é o mais prioritário dos 9 não deveria implicar, necessariamente, que seu peso é 8 vezes maior que o do segundo menos prioritário, e infinitamente maior que o do menos prioritários de todos.

A função escore utilizada para a classificação das áreas metropolitanas, norte-americanas é a média dos escores obtidos ponderada por fator. A classificação apresentada considera peso igual para todos os fatores.

A seguir apresentamos a classificação “abstrata” (independente das características dos sujeitos) e a classificação para alguns sujeitos específicos que responderam o inventário, com finalidade de mostrar como diferentes prioridades levam a diferentes classificações. A tabela a seguir, apresenta algumas características dos mesmos.

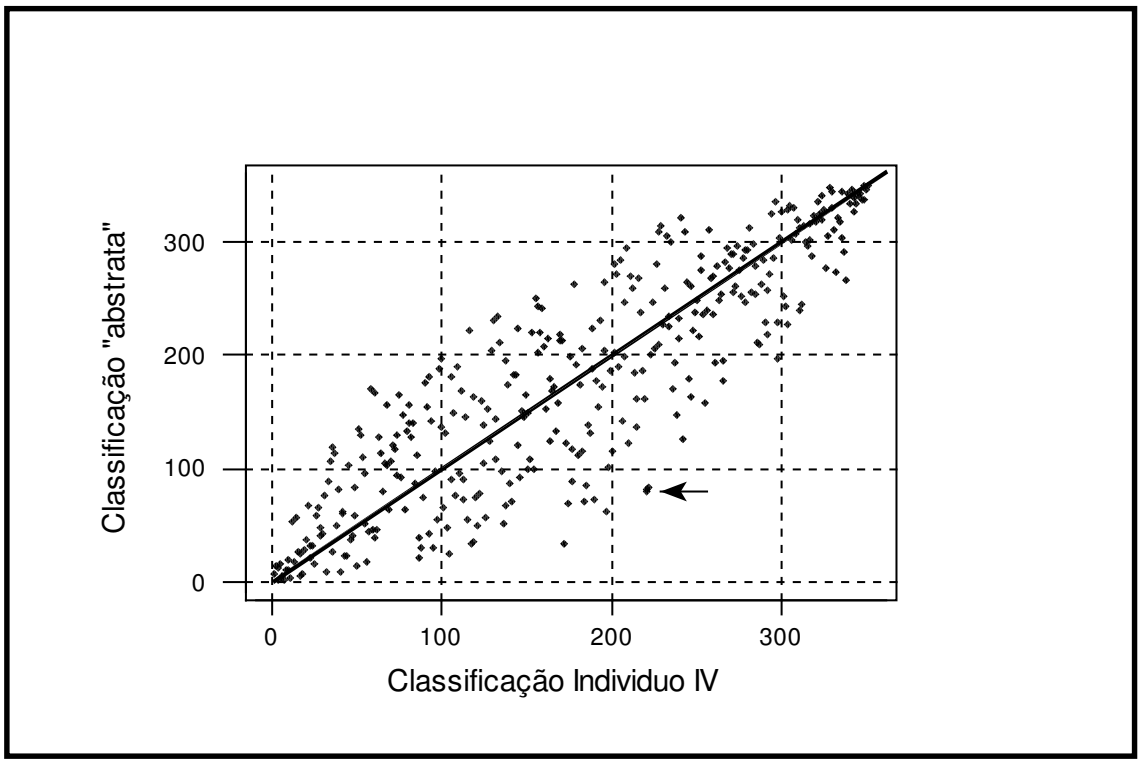
Tabela 2: Características dos sujeitos que responderam ao inventário

Sujeito	Sexo	Idade	Escolaridade			Estado Civil	Renda (R$)	Tem filhos?
			do sujeito	da mãe	do pai			
I	M	37	Superior	1º grau	1º grau	Casado	5000,00	Sim
II	F	71	Superior	1º grau	1º grau	Divorciada	7000,00	Sim
III	M	29	2º grau	1º grau	1º grau	Casado	1500,00	Sim
IV	F	35	1º grau	1º grau	1º grau	Divorciada	400,00	Sim
V	F	24	Superior	Superior	2º grau	Solteira	3500,00	Não
VI	M	53	Superior	2º grau	1º grau	Casado	10000,00	Sim

A seguir, apresentamos classificações resultantes de uma mesma função escore adotada (média ponderada dos fatores), porém com prioridades diversas para cada um dos fatores envolvidos.

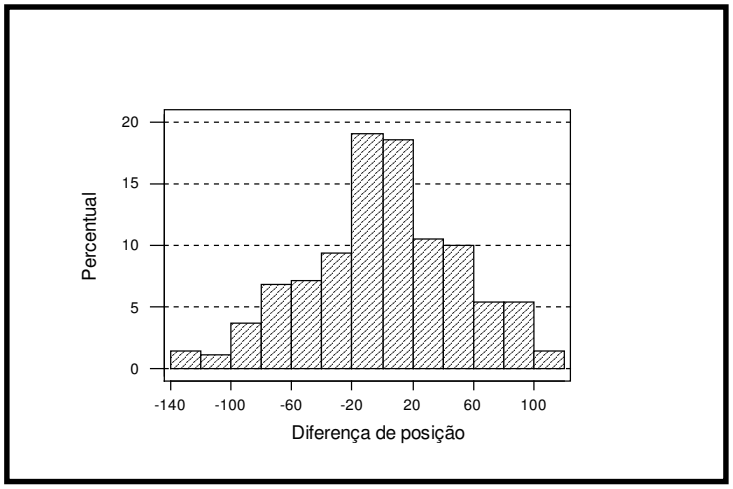
Veja na Figura 1, a seguir, que a classificação do indivíduo IV e a classificação abstrata são diferentes. Por exemplo, algumas áreas metropolitanas que na classificação abstrata receberam classificação melhor que a 100ª colocação, receberam, para os critérios do indivíduo IV uma classificação pior que 200ª (ver seta).

Gráfico 1: Classificação das áreas metropolitanas americanas segundo o Índice de Qualidade de Vida Urbana considerando um indivíduo específico e sem considerar os indivíduos



A seguir, no gráfico 2, mostramos o histograma das diferenças de posições obtida nas duas classificações. Pode-se ver que aproximadamente 65% das áreas metropolitanas diferem nas duas classificações em mais que 20 posições.

Gráfico 2: Distribuição das diferenças de posição obtidas nas duas classificações consideradas



Também apresentamos nas tabelas 3 e 4, os dez primeiros colocados para cada um dos 6 indivíduos, respectivamente. Veja que a concordância exata inexiste, mas a concordância com relação ao conjunto de áreas metropolitanas, entre os dez primeiros é razoável. Isto é, o conjunto de melhores cidades é bastante semelhante para eles, ainda que não concordem exatamente com a posição que a cidade se encontra na classificação.

Tabela 3: Classificação das 10 primeiras áreas metropolitanas americanas segundo o Índice de Qualidade de Vida Urbana para os indivíduos I, II e III entrevistados

Rank	Sujeito I	Sujeito II	Sujeito III
1	Houston, TX	Minneapolis - ST.Paul , MN – WI	Minneapolis - ST.Paul , MN – WI
2	Orange County, CA	Washington, DC - MD - VA - WV	Orange County, CA
3	Seattle - Bellevue - Everett, WA	Orange County, CA	Long Island, NY
4	Tampa - St. Petersburg - Clearwater, FL	Seattle - Bellevue - Everett, WA	Pittsburgh, PA
5	San Antonio, TX	Houston, TX	Toronto, ON
6	Phoenix - Mesa, AZ	Toronto, ON	Washington, DC - MD - VA – WV
7	Atlanta, GA	Atlanta, GA	Seattle - Bellevue - Everett, WA
8	San Diego, CA	Phoenix - Mesa, AZ	Philadelphia, PA - NJ
9	Washington, DC – MD – VA - WV	Philadelphia, PA - NJ	San Jose, CA
10	Riverside - San Bernardino, CA	Long Island, NY	Cincinnati, OH-KY-IN

Tabela 4: Classificação das 10 primeiras áreas metropolitanas americanas segundo o Índice de Qualidade de Vida Urbana para os indivíduos IV, V e VI entrevistados

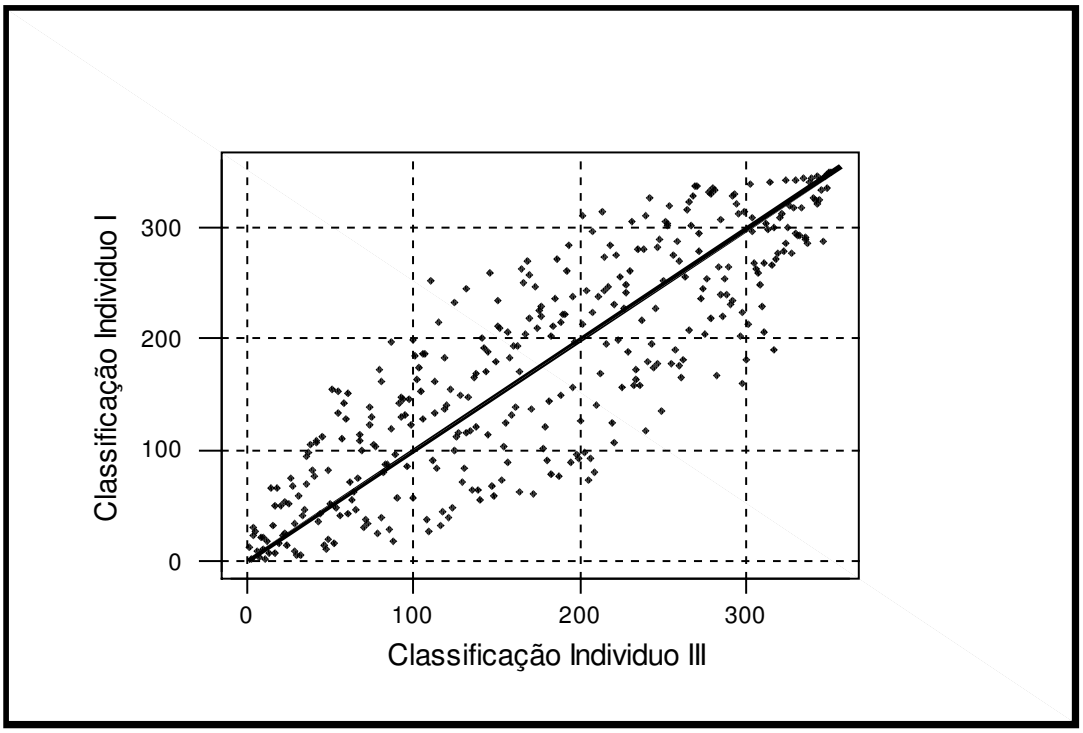
Rank	Sujeito IV	Sujeito V	Sujeito VI
1	Minneapolis - ST.Paul , MN - WI	Orange County, CA	Minneapolis - ST.Paul , MN - WI
2	Pittsburgh, PA	Seattle - Bellevue - Everett, WA	Orange County, CA
3	Long Island, NY	San Jose, CA	Pittsburgh, PA
4	Orange County, CA	Washington, DC - MD - VA - WV	Toronto, ON
5	Toronto, ON	Houston, TX	Washington, DC - MD - VA - WV
6	Washington, DC - MD - VA - WV	Minneapolis - ST.Paul , MN - WI	Seattle - Bellevue - Everett, WA
7	Seattle - Bellevue - Everett, WA	Long Island, NY	Long Island, NY
8	Philadelphia, PA - NJ	Toronto, ON	San Jose, CA
9	Cincinnati, OH-KY-IN	San Diego, CA	Philadelphia, PA - NJ
10	San Jose, CA	Phoenix - Mesa, AZ	Salt Lake City - Ogden, UT

Este comportamento onde um conjunto de cidades ocupa os primeiros lugares em diferentes classificações, se deve ao aspecto ***dominante*** que estas cidades apresentam.

Dizemos que uma cidade é dominante sobre outra quando, em todos os fatores analisados, ela apresenta escore maior que a outra, obtendo por conseqüência uma classificação melhor, que a da outra, independente do conjunto de pesos adotado. Isto se deve ao fato de estar sendo adotada uma função escore que considera pesos e variáveis maiores ou iguais a zero. Esta informação é importante na orientação de escolha de um sujeito pois, se ele concorda com os critérios adotados, então a cidade dominante – numa confrontação direta com a dominada – será sempre a escolha sobre uma cidade dominada.

No gráfico 3, a seguir, apresentamos as classificações para dois desses indivíduos. Pode-se ver a discordância na maioria das áreas metropolitanas, como já havíamos observado na comparação da classificação "abstrata" versus a classificação do individuo IV.

Gráfico 3: Classificação das áreas metropolitanas americanas segundo o Índice de Qualidade de Vida Urbana considerando os individuo I e III.



A função escore adotada foi a seguinte:

$$\mathscr{S} = a_1x_1 + a_2x_2 + a_3x_3 + \ldots + a_8x_8 + a_9x_9$$

Para a classificação "abstrata", independente de preferências, os $a_{i's} = 1$ e $x_{i's}$ são os escores combinados para cada um dos nove fatores, com cada escore variando de 0 (pior) a 100 (melhor). No caso, de cada indivíduo citado acima os $a_{i's}$ correspondem às suas preferências. A tabela a seguir, apresenta as preferências detectadas pelo inventário, por indivíduo.

Tabela 5: Peso de cada fator segundo os sujeitos que responderam ao inventário

Sujeito	**Fator**								
	Moradia	Transporte	Emprego	Educação	Clima	Segurança	Artes	Saúde	Recreação
I	15	7	21	10	14	11	4	14	4
II	7	11	19	7	6	13	11	19	7
III	10	10	17	15	3	22	3	18	3
IV	13	8	15	17	3	22	4	14	4
V	6	13	18	4	14	19	7	13	7
VI	8	19	10	10	7	18	13	3	13

Podemos ver através das preferências apresentadas que para estes indivíduos, a segurança e/ou o emprego são prioridades. A classificação de áreas metropolitanas sem considerar o sujeito, provavelmente, decorre da dificuldade em quantificar as preferências do mesmo. Responder qualitativamente, por exemplo, que segurança e oportunidades de emprego são prioridades, não é uma tarefa tão difícil quanto definir o grau de importância relativa dos mesmos.

Já que o peso dado para cada variável determina a sua colocação, uma situação interessante é saber com quais pesos (prioridades) determinada área metropolitana ocuparia o primeiro lugar (ou o último).

Para obtermos as classificações acima, em nenhum momento foi apresentado o nome de qualquer cidade para os sujeitos. Assim Houston-TX foi considerada a melhor

cidade para o sujeito I, baseado única e exclusivamente nas características da cidade, e nas prioridades relativas do sujeito para cada uma destas características.

Vejamos, por exemplo, se existem pesos que levariam Nova Iorque ao primeiro lugar. E, se existem, quais são eles?

Para responder estas questão, seja M a matriz dos escores (355 x 9) e $M_{\text{Nova Iorque}}$ a linha correspondente a Nova Iorque. Construímos

$$M' = (1, 1, 1, \ldots, 1)^t \cdot M_{\text{Nova Iorque}} - M$$

onde a i-ésima linha de M' é a diferença entre $M_{\text{Nova Iorque}}$ e a i-ésima linha de M. Nós podemos encontrar o vetor de pesos $\underline{a}$ tal que $M'a \geq 0$, onde $a_i \geq 0$ e $\Sigma\, a_i = 1$. Isto significa que, dado os pesos $\underline{a}$, a soma dos escores ponderados por estes pesos para Nova Iorque é maior que a soma dos escores ponderados por estes pesos para qualquer outra cidade.

Resolver este problema é equivalente a encontrar uma solução factível para um problema de programação linear. Neste exemplo, especificamente, uma solução em termos percentuais para o vetor de prioridades é, 15% para Facilidade de Transporte, 20% para Sistema de Educação, 30% para as Opções Culturais, 25% para os Sistemas de Saúde, 10% para as Opções de Lazer e 0% para Oportunidades de Emprego, custo de moradia, Características do Clima e Segurança. Esta é uma solução factível, porém não é a única.

RESUMO

A classificação de áreas metropolitanas segundo um Índice de Qualidade de Vida Urbana, envolve as seguintes etapas:

1. Definição dos critérios que irão medir a Qualidade de Vida Urbana;
2. Definição dos indicadores para cada um dos critérios;
3. A forma como medi-los e
4. A definição da função escore (função matemática) que irá combinar cada um dos critérios considerados, segundo as prioridades dos sujeitos, para a construção da classificação (ranking).

Situação Atual = A função escore considera somente as características dos objetos, deixando implícito que a classificação independe do sujeito.

A QUESTÃO DOS VESTIBULARES

Num certo sentido os exames vestibulares das grandes universidades brasileiras já praticam, de forma um tanto rudimentar, a abordagem bipolar para a classificação de objetos (os alunos) para as suas diversas faculdades (os sujeitos). Com uma abordagem multicritérios, adota-se uma ponderação por disciplina que varia de curso para curso. A UNICAMP tem adotado, na segunda fase do seu vestibular, desde 1995, um critério de ponderação que dá peso 2 para as disciplinas consideradas prioritárias. Na tabela abaixo vemos estes pesos para 3 faculdades: Engenharia Elétrica, Medicina e Economia.

Tabela 6: Conjunto de pesos adotado no Vestibular da UNICAMP 99, por curso

Curso	1ª. Fase	Português	Matemática	Química	Biologia	Física	História	Geografia	Língua Estrangeira
Engenharia Elétrica	2	1	2	1	1	2	1	1	1
Medicina	2	1	1	2	2	1	1	1	1
Economia	2	1	2	1	1	1	2	1	1

Embora se constitua numa evolução no sentido da abordagem bipolar, a distribuição de pesos é ainda tosca e arbitrária. Ela adota, intrinsecamente, uma classificação qualitativa binária de disciplinas – prioritárias e não prioritárias – e uma distribuição mais ou menos arbitrária de pesos: 2 para as prioritárias e 1 para as não prioritárias. As conseqüências destas escolhas são sérias, com impacto decisivo no resultado final: a classificação – e conseqüente seleção – dos candidatos, conforme veremos a seguir.

È importante rastrear, também para este caso, todas as etapas envolvidas no processo de classificação, revelando sua analogia com o caso da classificação segundo um índice de qualidade de vida:

1. **Definição dos critérios.** Consiste no programa do ensino médio adotado pelo MEC, para cada uma das disciplinas que compõem o vestibular (na UNICAMP é Matemática, Física, História, Geografia, Biologia, Química, Língua Estrangeira, Língua Portuguesa e uma Redação) e publicado no Manual do Candidato. Como no caso do índice de qualidade de vida urbana, é nesta etapa que se concentram as principais discussões. Em particular, ressaltamos o debate sobre a habilidade do critério Redação de medir – de maneira objetiva e isenta – a capacidade de raciocínio e concatenação de idéias por parte do vestibulando [4] – [7], [12]-[14], [22], [35], [40], [41], [46] - [50] ;
2. **O instrumento de medição dos critérios**. Baseado nestes critérios é realizada, na primeira fase, uma prova geral e uma redação e, na segunda fase, provas dissertativas em cada uma das disciplinas mencionadas.
3. **A função escore utilizada**. A função escore adotada é a soma total de pontos para os vestibulandos que foram aprovados na primeira fase. O número de alunos classificados – por curso – na primeira fase é limitado acima a 25 vezes o número de vagas. A função escore adotada encontra-se abaixo.

$$\mathscr{S} = a_1 Z_1 + a_2 Z_2 + a_3 Z_3 + a_4 Z_4 + \ldots + a_9 Z_9$$

Onde os pesos a_i variam para cada disciplina, por faculdade, como vimos anteriormente as variáveis Z são transformações lineares das variáveis X correspondentes, para padronização com média 500 e desvio padrão 100:

$$Z_i = \left(\frac{X_i - \mu_X}{\sigma_X} \right) 100 + 500$$

onde μ_X é a média das notas dos alunos na disciplina X e σ_x é o desvio padrão das notas dos alunos na disciplina X.

Assim, para se obter a pontuação – o escore – de um vestibulando para uma determinada prova (disciplina), necessitamos da sua nota nesta disciplina, a nota média obtida por todos os alunos nesta disciplina e o desvio padrão das notas de todos os alunos nesta disciplina. A padronização permite considerar igualmente provas que tenham graus de dificuldade diferentes.

Considerem-se aqui os resultados dos vestibulares da UNICAMP de 1999, para aqueles três cursos. O número de candidatos na segunda fase foi de 652, 2756 e 542 para Engenharia Elétrica, Medicina e Economia, respectivamente. Os candidatos são identificados por um número que corresponde às suas classificações segundo os pesos adotados. Vejamos o que aconteceria, se adotássemos outro conjunto de pesos, por exemplo:

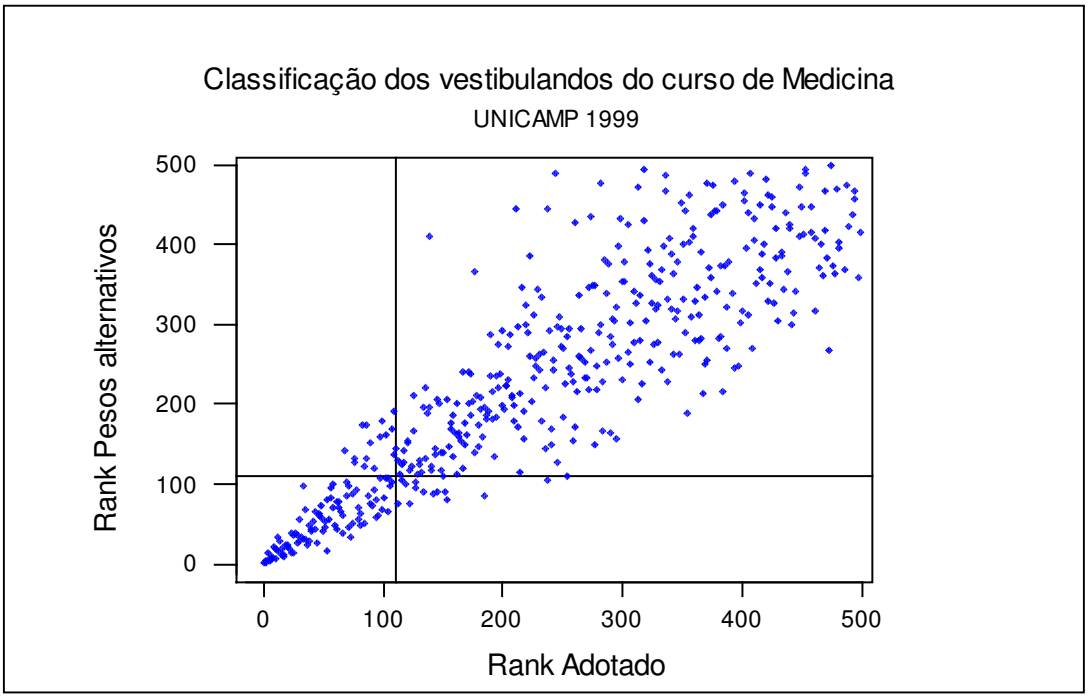
Tabela 7: Conjunto alternativo de pesos, por curso

Curso	1ª. Fase	Português	Matemática	Química	Biologia	Física	História	Geografia	Língua Estrangeira
Engenharia Elétrica	1	1	2,5	1	1	2,5	1	0,5	1,5
Medicina	1	1	1	2	2,5	1	1	1	1,5
Economia	1	1,5	2,5	1	1	1	2	1	1

No caso da Medicina, que oferece 110 vagas por ano, aumentamos o peso de Biologia e de Língua Estrangeira (poderia ser justificado pela grande quantidade de bibliografia em inglês) e diminuímos o peso da primeira fase, por entender que ela já teve um caráter absoluto, de pré-classificação. Não estamos propondo seriamente um novo conjunto de pesos para Medicina, mas apenas alertando para a sensibilidade do resultado final com relação à escolha destes pesos. Na Figura 4 relacionamos a classificação dos 500 melhores alunos – segundo o critério usual – com relação ao critério usual e ao considerado acima.

Além de gerar outra classificação, mudaria a condição de 15 vagas de aprovado para reprovado e vice-versa, isto é, 14% das vagas seriam substituídas por outros alunos.

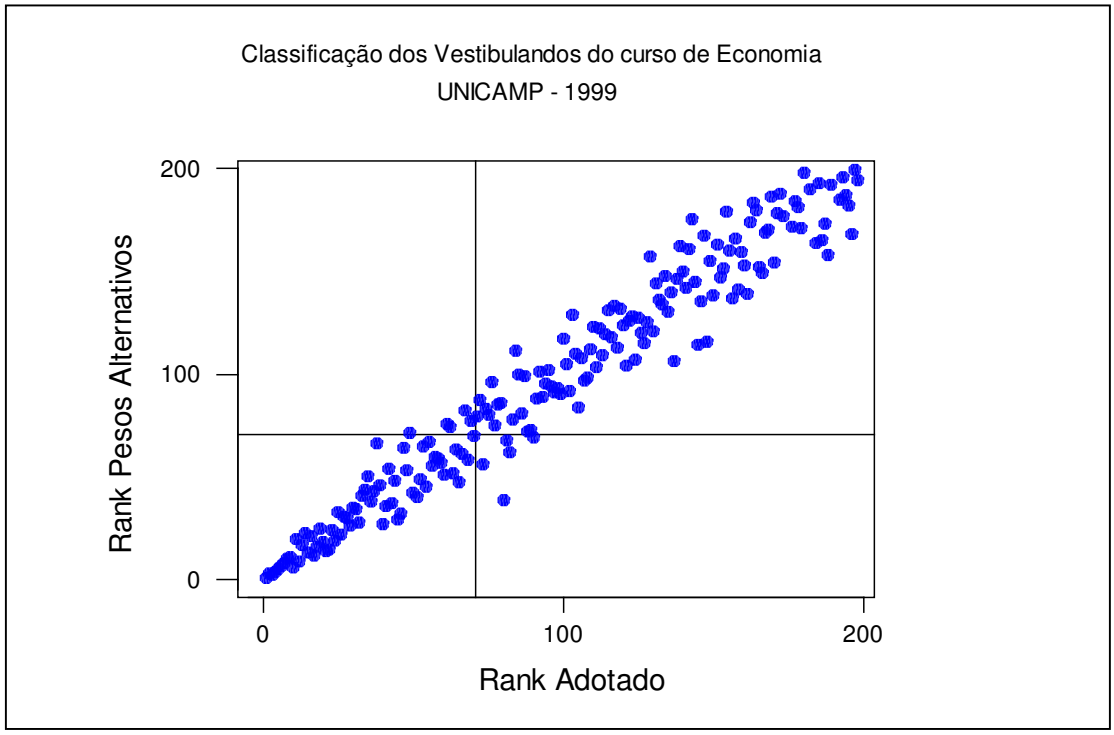
Gráfico 4: Classificação dos vestibulandos do curso de Medicina – UNICAMP 1999 – segundo o conjunto de pesos adotado atualmente e um conjunto de pesos alternativo



No gráfico a seguir vemos a situação no curso de Economia, com 70 vagas anuais, sendo oferecida para o curso diurno, as mudanças para o vetor de pesos sugerido, são menores, mas existem, neste caso 5 alunos que trocariam a situação de aprovação por reprovação. O fato é que pequenas alterações nas ponderações dos critérios (provas) influenciam a classificação dos alunos e conseqüentemente sua aprovação.

No caso do vestibular também observamos que alguns alunos povoam os primeiros lugares (e os últimos) independente do conjunto de pesos adotados. Ocorre aqui o fenômeno de dominância. Um aluno que se apresentou melhor que outro em todas as provas realizadas, terá uma classificação melhor que o outro, independente do vetor de pesos adotados. Dizemos que ocorre entre os dois uma relação de dominador/dominado.

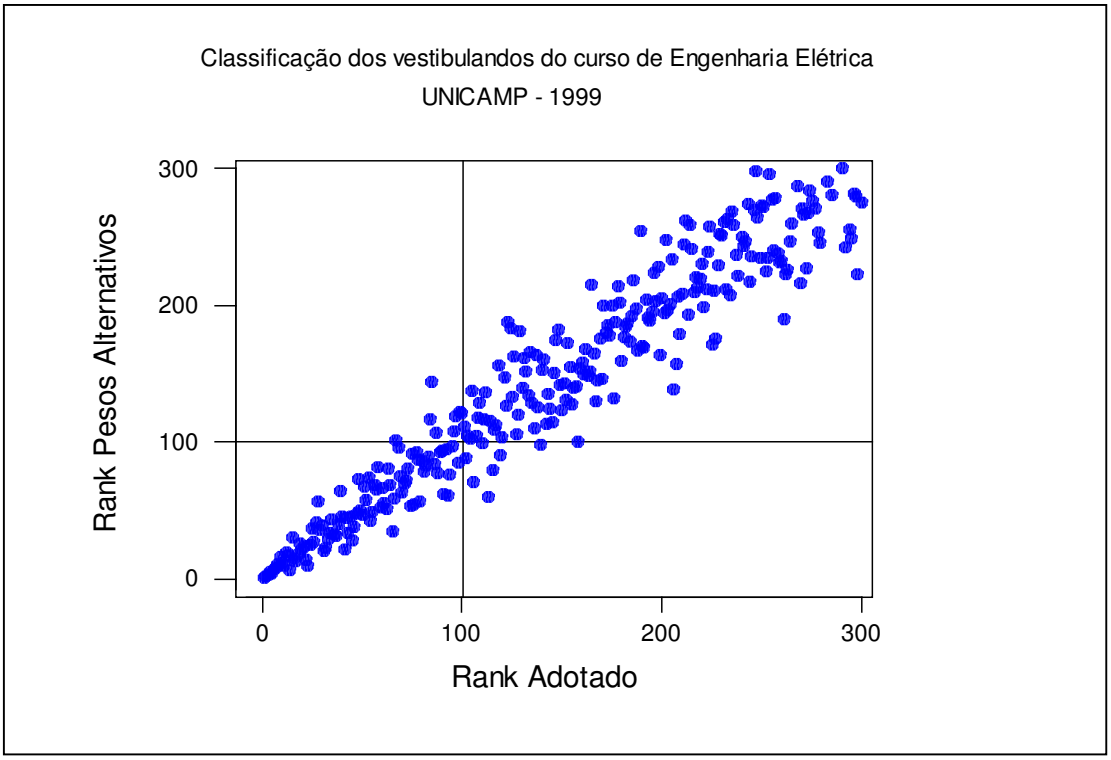
Gráfico 5: Classificação dos vestibulandos do curso de Economia – UNICAMP 1999 – segundo o conjunto de pesos adotado atualmente e um conjunto de pesos alternativo



Para o curso de Engenharia Elétrica, que oferece 100 vagas, distribuídas entre o diurno (70) e o noturno (30), a situação é análoga, como vemos no gráfico 6, a seguir. No conjunto de pesos sugerido, aumentamos a importância na pontuação de Matemática, Física e Língua Estrangeira e diminuímos a importância de História, uma de muitas situações possíveis – e razoavelmente defensáveis – para as prioridades na escolha dos alunos do referido curso.

A relevância destes exemplos, de classificação, deve-se ao esquecimento generalizado, ou pequena importância ao seu caráter bipolar e a função escore utilizada para gerar a classificação. Veja que no caso das áreas metropolitanas, as discussões limitam-se aos critérios a serem utilizados para classificação, deixando a função escore para uma discussão secundária. No caso do vestibular os critérios são previamente discutidos através do programa do MEC, a bipolaridade é superficialmente colocada quando adotamos pesos diferentes para cursos diferentes e a função escore (que resulta na soma total dos pontos) é arbitrariamente definida quando define peso 2 para as matérias prioritárias e peso 1 para as demais.

Gráfico 6: Classificação dos vestibulandos do curso de Engenharia Elétrica – UNICAMP 1999 - segundo o conjunto de pesos adotado atualmente e um conjunto de pesos alternativo



É importante ressaltar que no caso do vestibular, a função escore foi definida quantitativamente baseada numa informação qualitativa (matérias prioritárias), isto é, em nenhum momento se questionou o quão prioritário era uma determinada matéria e a sua relação de barganha com as outras. Em outras palavras, não se discutiu quantos pontos um vestibulando de Engenharia Elétrica precisa fazer a mais na prova de Biologia para compensar cada ponto a menos que ele fizer na prova de Física, por exemplo.

Veremos, mais adiante, como resolvemos este problema em detalhes. Genericamente falando, o que sugerimos e mostramos rigorosamente é que observando como classificam os sujeitos os respectivos objetos, encontramos a função escore geral que os orienta. A relevância disto deve-se ao fato que os sujeitos na sua totalidade (ou quase totalidade) desconhecem a sua própria lógica de classificação, a qual é filtrada e revelada pelas suas escolhas.

CAPÍTULO II -ESTIMAÇÃO POR MÍNIMOS QUADRADOS

APRESENTAÇÃO

Em diversos problemas de classificação e escolha, é razoável se admitir que, confrontado com um objeto, qualquer sujeito pode oferecer uma avaliação objetiva, embora imperfeita, do seu escore para aquele objeto. Seja o modelo

$$\mathscr{Y} = \mathscr{Y}(\boldsymbol{s}, \boldsymbol{o}) = \mathscr{S}\left[\underset{\sim}{\boldsymbol{w}}(\boldsymbol{s}),\ \underset{\sim}{\boldsymbol{x}}(\boldsymbol{o})\right] + \delta$$

Como já mencionamos na Introdução Geral, $\mathscr{S}\left[\underset{\sim}{\boldsymbol{w}}(\boldsymbol{s}),\ \underset{\sim}{\boldsymbol{x}}(\boldsymbol{o})\right]$ é o escore geral dado por sujeitos com as características $\underset{\sim}{w}(s)$ para objetos com as características $\underset{\sim}{x}(o)$ e δ é a componente especifica de um sujeito para um objeto. Assim dois sujeitos idênticos em termos $\underset{\sim}{w}(s)$ avaliando dois objetos idênticos em termos de $\underset{\sim}{x}(o)$, apresentarão o mesmo escore geral ($\mathscr{S}\left[\underset{\sim}{\boldsymbol{w}}(\boldsymbol{s}),\ \underset{\sim}{\boldsymbol{x}}(\boldsymbol{o})\right]$), diferenciando cada escore na sua componente específica δ.

Seja y a avaliação de **o** dada por **s.** Temos então

$$y = \mathscr{S}(\boldsymbol{s}, \boldsymbol{o}) + \tau = \mathscr{S}\left[\underset{\sim}{\boldsymbol{w}}(\boldsymbol{s}),\ \underset{\sim}{\boldsymbol{x}}(\boldsymbol{o})\right] + \delta + \tau$$

onde τ é a imprecisão associada a avaliação do sujeito $\boldsymbol{s}$ para o objeto $\boldsymbol{o}$, assumiremos que τ tem distribuição normal, com esperança zero e variância constante σ^2.

Como o nosso interesse é o de construir modelos para Y, a especificidade δ será agregada a τ, num único erro experimental ε. Assim,

$$y = \mathscr{Y}(\boldsymbol{s}, \boldsymbol{o}) + \varepsilon = \mathscr{S}\left[\underset{\sim}{\boldsymbol{w}}(\boldsymbol{s}),\ \underset{\sim}{\boldsymbol{x}}(\boldsymbol{o})\right] + \varepsilon$$

UM EXEMPLO NUMÉRICO

Para fixar os conceitos apresentados, analisaremos um exemplo numérico simples. Temos 100 objetos, definidos por duas características (x_1 , x_2) com valores entre 0 e 1. Estes objetos serão classificados por um grupo de 100 sujeitos, especificados por suas características (w_1 , w_2), também com valores entre 0 e 1. As tabelas a seguir apresentam os valores numéricos das características dos objetos e dos sujeitos. Os valores foram gerados com distribuição uniforme no intervalo 0 a 1.

Tabela 8: Apresentação dos objetos e sujeitos segundo os valores numéricos de suas características

I	Objeto		Sujeito		I	Objeto		Sujeito		I	Objeto		Sujeito	
	X_1	X_2	W_1	W_2		X_1	X_2	W_1	W_2		X_1	X_2	W_1	W_2
1	0,09	0,40	0,63	0,38	**35**	0,04	0,70	0,67	0,66	**69**	0,87	0,86	0,20	0,67
2	0,18	0,03	0,49	0,41	**36**	0,11	0,44	0,57	0,40	**70**	0,51	0,01	0,53	0,40
3	0,20	0,82	0,12	0,50	**37**	0,11	0,24	0,42	0,40	**71**	0,32	0,65	0,61	0,61
4	0,57	0,14	0,44	0,62	**38**	0,79	0,37	0,23	0,63	**72**	0,28	0,81	0,70	0,78
5	0,29	0,19	0,65	0,35	**39**	0,04	0,43	0,46	0,67	**73**	0,59	0,68	0,07	0,45
6	0,91	0,96	0,78	0,36	**40**	0,93	0,98	0,73	0,80	**74**	0,75	0,89	0,38	0,64
7	0,99	0,83	0,68	0,57	**41**	1,00	0,68	0,91	0,44	**75**	0,15	0,44	0,53	0,41
8	0,42	0,58	0,42	0,28	**42**	0,56	0,17	0,42	0,59	**76**	0,08	0,91	0,66	0,96
9	0,02	0,77	0,27	0,50	**43**	0,25	0,13	0,74	0,47	**77**	0,62	0,00	0,53	0,76
10	0,64	0,46	0,39	0,57	**44**	0,98	0,17	0,40	0,66	**78**	0,10	0,17	0,32	0,50
11	0.64	0.40	0,94	0,63	**45**	0,84	0,87	0,27	0,59	**79**	0,55	0,76	0,42	0,72
12	0,77	0,54	0,60	0,22	**46**	0,59	0,23	0,69	0,62	**80**	0,64	0,56	0,67	0,46
13	0,47	0,73	0,00	0,58	**47**	0,26	0,55	0,39	0,59	**81**	0,88	0,23	0,51	0,81
14	0,58	0,59	0,51	0,49	**48**	0,98	0,25	0,42	0,49	**82**	0,53	0,20	0,86	0,49
15	0,19	0,97	0,40	0,69	**49**	0,01	0,30	0,51	0,59	**83**	0,96	0,97	0,35	0,75
16	0,60	0,81	0,34	0,22	**50**	0,79	0,28	0,55	0,54	**84**	0,10	0,91	1,00	0,54
17	0,05	0,33	0,61	0,06	**51**	0,52	0,95	0,73	0,19	**85**	0,75	0,04	0,28	0,25
18	0,30	0,75	0,26	0,29	**52**	0,11	0,49	0,43	0,51	**86**	0,72	0,20	0,46	0,47
19	0,27	0,84	0,71	0,52	**53**	0,19	0,52	0,43	0,70	**87**	0,25	0,76	0,56	0,05
20	0,48	0,81	0,50	0,19	**54**	0,18	0,09	0,75	0,79	**88**	0,22	0,29	0,94	0,27
21	0,57	0,40	0,38	0,55	**55**	0,41	0,99	0,49	0,89	**89**	0,21	0,80	0,58	0,35
22	0.61	0.58	0,18	0,42	**56**	0,75	0,74	0,49	0,72	**90**	0,09	0,28	0,55	0,53
23	0,63	0,54	0,47	0,19	**57**	0,15	0,76	0,29	1,00	**91**	0,25	0,12	0,53	0,20
24	0,07	0,54	0,56	0,34	**58**	0,89	0,95	0,53	0,31	**92**	0,14	0,46	0,53	0,62
25	0,86	0,31	0,43	0,62	**59**	0,49	0,19	0,35	0,55	**93**	0,28	0,35	0,52	0,40
26	0,49	0,14	0,47	0,45	**60**	0,69	0,94	0,40	0,56	**94**	0,55	0,08	0,38	0,30
27	0,39	0,10	0,14	0,53	**61**	0,40	0,79	0,24	0,00	**95**	0,97	0,47	0,57	0,58
28	0,35	0,34	0,57	0,58	**62**	0,62	0,47	0,36	0,29	**96**	0,57	0,99	0,56	0,28
29	0,57	0,34	0,50	0,66	**63**	0,73	0,42	0,48	0,67	**97**	0,47	0,41	0,80	0,50
30	0,65	0,78	0,53	0,61	**64**	0,39	0,27	0,68	0,44	**98**	0,60	0,01	0,63	0,41
31	0,23	0,46	0,73	0,42	**65**	0,96	0,76	0,70	0,54	**99**	0,12	0,35	0,33	0,35
32	0,20	0,30	0,48	0,69	**66**	0,15	0,89	0,48	0,61	**100**	0,69	0,73	0,81	0,20
33	0.99	0.04	0,61	0,65	**67**	1,00	0,77	0,29	0,36					
34	0,89	0,70	0,39	0,50	**68**	0,42	0,33	0,43	0,68					

Vamos classificar, inicialmente, os objetos segundo a abordagem clássica, considerando exclusivamente as características dos objetos. Poderíamos adotar qualquer

vetor de pesos para obtermos um escore para cada objeto, adotaremos pesos iguais (procedimento mais comum) neste exemplo.

Desta forma, o escore obtido para cada objeto é a média aritmética dos valores numéricos das duas características. À classificação induzida por estes escores denominaremos *rank abstrato*. Na tabela abaixo apresentamos, como ilustração, os escores médios e a respectiva classificação para os 10 primeiros e os 10 últimos objetos apresentados na tabela anterior.

Tabela 9: Apresentação dos Escores Médios para os objetos e sua respectiva classificação.

Objeto	X_1	X_2	Escore	Rank
1	0.0920	0.3980	0.2450	**87**
2	0.1790	0.0335	0.1063	**100**
3	0.1954	0.8197	0.5076	**48**
4	0.5721	0.1351	0.3536	**69**
5	0.2880	0.1905	0.2393	**89**
6	0.9104	0.9646	0.9375	**3**
7	0.9942	0.8263	0.9103	**5**
8	0.4162	0.5767	0.4965	**52**
9	0.0222	0.7749	0.3986	**62**
10	0.6372	0.4554	0.5463	**40**
. . .				
83	0.9563	0.9693	0.9628	**1**

Objeto	X_1	X_2	Escore	Rank
84	0.1006	0.9133	0.5070	**49**
. . .				
91	0.2486	0.1237	0.1862	**93**
92	0.1410	0.4589	0.3000	**81**
93	0.2837	0.3520	0.3179	**75**
94	0.5468	0.0793	0.3131	**76**
95	0.9731	0.4723	0.7227	**17**
96	0.5698	0.9861	0.7780	**14**
97	0.4685	0.4078	0.4382	**59**
98	0.6000	0.0141	0.3071	**78**
99	0.1236	0.3507	0.2372	**90**
100	0.6929	0.7281	0.7105	**19**

Na tabela acima vemos que o *objeto 83*, baseado no escore médio, foi classificado em primeiro lugar e o *objeto 2*, foi classificado em centésimo (último) lugar e, assim por diante. Esta abordagem do problema da classificação dos objetos considera, implicitamente que:

- ✓ as características dos objetos reúnem em si, todas as informações disponíveis para sua ordenação;
- ✓ Os pesos das diferentes características na formação do escore são iguais; e
- ✓ as preferências dos sujeitos com relação às características do objeto são constantes em $\boldsymbol{S}$, para qualquer objeto $\boldsymbol{o} \in \boldsymbol{O}$.

NO SISTEMA BIPOLAR DE CLASSIFICAÇÃO, que estamos introduzindo, o escore não é mais uma característica própria dos objetos, independente dos sujeitos, mas ganha agora o caráter de uma propriedade do par (***s, o***) segundo um modelo como definimos anteriormente. Estamos considerando que as preferências dos sujeitos com relação às características do objeto não são mais constantes em ***S***.

Na construção de modelos bipolares para o escore, assumiremos que:

- ✓ os escores dos objetos dependem também das características dos sujeitos;
- ✓ a influência das diferentes variáveis na formação do escore, é governada por funções matemáticas mais complexas que simples médias aritméticas ou médias ponderadas das mesmas;

e neste trabalho consideraremos sempre que:

- ✓ a função que associa a um par (***s, o***) um escore real, pode ser adequadamente modelada por funções polinomiais de $\underset{\sim}{w}$ e $\underset{\sim}{x}$, de grau não superior a 2.

Admitimos aqui a hipótese central da Análise de Regressão: a da existência de um modelo determinístico implícito conectando variáveis explicativas e respostas. No nosso caso, esta hipótese equivale à aceitação da existência de uma função escore implícita no processo de escolha.

Estas funções serão modeladas por polinômios de grau não superior a dois. Os parâmetros destas funções polinomiais serão estimados a partir de dados experimentais. Neste capítulo tratamos das situações em que confrontado com um objeto qualquer $o \in O$, qualquer sujeito $s \in S$ pode fornecer uma avaliação objetiva, embora imperfeita, do seu escore absoluto para aquele objeto.

Conforme veremos, em grandes levantamentos amostrais, os modelos estimados freqüentemente induzirão uma ordenação do ***O*** segundo as preferências de qualquer sujeito $s \in S$, muito mais precisa que aquela produzida pelo próprio sujeito ***s***. A qualidade relativa

da ordenação via modelo estimado e aquela produzida diretamente pelo sujeito, dependerá da relação entre a variância do erro de avaliação τ, e da variância da especificidade δ.

Pequenos valores de $V(\delta)$ geralmente estão associados a alto poder preditivo do modelo estimado quanto ao escore do par (**s, o**). De qualquer forma, nos casos de abundante disponibilidade de dados experimentais, o modelo associará, com grande precisão, o escore médio de um grupo homogêneo de sujeitos para um grupo homogêneo de objetos.

Vamos ilustrar estas idéias, através de simulação, lembrando que estamos adotando a modelação básica descrita anteriormente, para função escore:

$$\mathscr{S}\left[\underset{\sim}{\boldsymbol{w}}(\boldsymbol{s}),\ \underset{\sim}{\boldsymbol{x}}(\boldsymbol{o})\right] = \alpha_o+\alpha_1 x_1+\alpha_2 x_2+\alpha_{12} x_1 x_2+\alpha_{11} x_1^2+\alpha_{22} x_2^2$$

com $\alpha_i = \beta_{i,0}+ \beta_{i,1} w_1+ \beta_{i,2} w_2$, para todo i. Como já salientamos, este modelo envolve um total de 18 parâmetros.

Na construção de um caso para simulação, com valores arbitrários, vamos considerar que:

1) A função escore, é representada por uma função polinomial da família descrita anteriormente, com parâmetros dados pela matriz:

$$\beta = \begin{bmatrix} \beta_{0,0} & \beta_{0,1} & \beta_{0,2} \\ \beta_{1,0} & \beta_{1,1} & \beta_{1,2} \\ \beta_{2,0} & \beta_{2,1} & \beta_{2,2} \\ \beta_{11,0} & \beta_{11,1} & \beta_{11,2} \\ \beta_{12,0} & \beta_{12,1} & \beta_{12,2} \\ \beta_{22,0} & \beta_{22,1} & \beta_{22,2} \end{bmatrix} = \begin{bmatrix} 60.0 & 2.5 & -4.0 \\ 16.0 & -1.5 & 2.8 \\ 28.0 & -4.0 & 6.0 \\ -4.5 & 12.0 & -8.0 \\ 4.0 & 5.0 & -7.0 \\ -5.8 & -10.0 & 12.0 \end{bmatrix}$$

2) Sorteamos uma amostra aleatória, com reposição de 1000 pares de (***s, o***) .

3) Calculamos o escore obtido para cada par (objeto, sujeito) segundo a expressão:

$$\mathscr{S}\left[\underset{\sim}{\boldsymbol{w}}(\boldsymbol{s}),\ \underset{\sim}{\boldsymbol{x}}(\boldsymbol{o})\right]=\alpha_o+\alpha_1x_1+\alpha_2x_2+\alpha_{12}x_1x_2+\alpha_{11}x_1^2+\alpha_{22}x_2^2$$

com $\alpha_i = \beta_{i,0}+ \beta_{i,1}w_1+ \beta_{i,2}w_2$, para todo i. Chamaremos este escore de ***escore real***, num caso real impossível de ser acessado, devido ao desconhecimento da matriz β.

Introduzimos um desvio aleatório ε no ***escore real*** (resultante da agregação da especificidade à imprecisão do observador, definidas anteriormente), com distribuição normal de média zero e desvio padrão igual cinco (valor arbitrário), lembrando que o desvio padrão para os dados gerados foi igual a um. Ao valor resultante denominaremos ***escore observado***, representado por ***y.***

4) Usando os pares (***s, o***) e o ***escore observado*** estimamos pelo método dos Mínimos Quadrados a matriz de parâmetros $\boldsymbol{\beta}$, que denominaremos $\hat{\beta}$ e

5) Usando $\hat{\beta}$ temos, para cada sujeito, os escores dos objetos, e conseqüentemente a sua classificação.

Baseando-se nas escolhas feitas pelos sujeitos em geral, estimamos a função escore geral. Aplicando a mesma as características do sujeito e as características dos objetos, temos o ***rank estimado*** dos objetos para aquele tipo de sujeito especificamente. No gráfico a seguir podemos ver o rank observado x rank real e o rank estimado x rank real, ao mesmo tempo. Veja que o rank estimado se parece mais com o rank real que o rank observado. Ou seja, o modelo depura as imperfeições cometidas por cada tipo de sujeito, estimando uma função escore geral, desconhecida, mas que esta implícita no processo de escolha de cada sujeito.

Outro detalhe importante, é que a função escore estimada utilizou informações de vários indivíduos distintos, que permitiram gerar uma classificação melhor que a observada para cada individuo especificamente. As escolhas dos outros indivíduos permitiram a estimação da matriz de ***Betas*** , a qual define a função escore geral e que, permite a definição das funções escores específicas de cada individuo, dependendo do seu vetor de características $\underset{\sim}{w}(x)$.

Nesta simulação a Matriz ***Beta*** estimada foi $\hat{\beta} = \begin{bmatrix} 59.0 & 14.5 & -18.2 \\ 7.5 & 0.6 & 16.2 \\ 21.2 & -3.7 & 18.2 \\ 5.7 & -4.5 & -4.9 \\ 8.5 & 11.9 & -20.3 \\ -0.1 & -16.6 & -11.6 \end{bmatrix}$ que define a função escore geral e aplicada, por exemplo, ao vetor de características do indivíduo 1, $\underset{\sim}{s} = (0.63\,;\,0.38)$ temos o vetor de alfas estimados para o indivíduo 1 que é igual a $\underset{\sim}{\hat{\alpha}} = (61.2\,;14.1\,;25.8\,;1.0\,;8.3\,;-14.9)$ e define a função escore específica do individuo 1. Ao aplicarmos esta função escore aos objetos apresentados na tabela 8, temos o escore estimado para cada um dos 100 objetos em relação ao individuo 1.

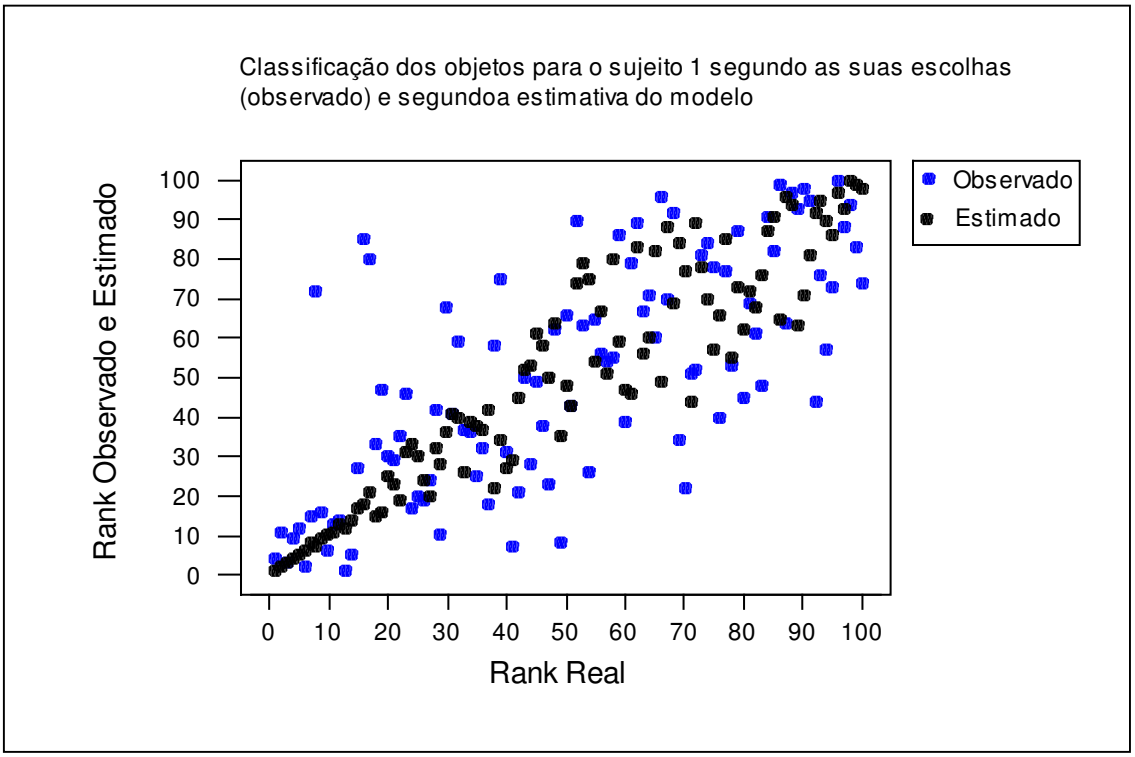
O escore real para cada um dos 100 objetos em relação ao individuo 1, é obtido aplicando a Matriz ***Beta*** ao vetor de características do individuo 1, resultando na função escore real para o individuo 1 que é igual a $\underset{\sim}{\hat{\alpha}} = (60.1\,;16.1\,;27.8\,;0.02\,;4.5\,;-1.9)$.

Os escore observados foram obtidos somando aos escores reais um erro com distribuição normal de média zero e desvio padrão cinco, como mencionamos anteriormente.

Assim temos para cada um dos 100 objetos seu ***escore real***, ***escore observado*** e ***escore estimado***, são eles que definem a ***classificação real*** (usaremos ***rank real***), ***classificação observada*** e ***classificação estimada***.

A seguir, apresentamos o comportamento das classificações observada x real e estimada x real , para o indivíduo 1.

Gráfico 7: Classificação dos objetos para o individuo 1 segundo as suas observações e a estimativa do modelo, considerando o desvio padrão = 5, para a soma da variabilidade da especificidade e imprecisão



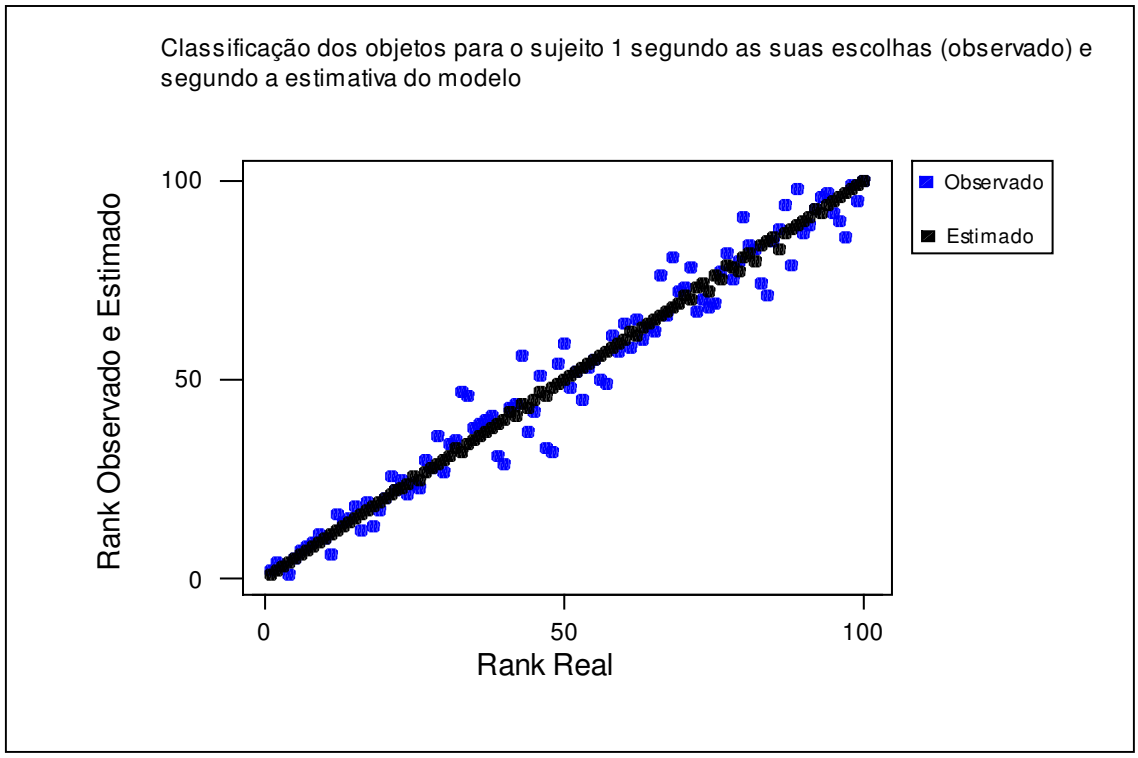
É importante verificar o caráter filtrador do Método de Estimação por Mínimos Quadrados da função escore. Veja que a função escore estimada gera uma classificação mais próxima da classificação real que a classificação observada (os pontos pretos acima estão mais concentrados que os azuis, em torno da reta imaginária, $y=x$).

Resumindo, na simulação acima foram apresentados objetos com características distintas para indivíduos distintos, com a finalidade de conhecer como eles, segundo suas características, classificam estes objetos. Introduzimos uma imperfeição na avaliação de cada sujeito para cada objeto, fruto da imprecisão em avaliar somado a especificidade de cada par (sujeito,objeto). A função escore geral estimada permite a construção das funções escores específicas para cada individuo e, nos revela que a função escore estimada específica para um individuo, utilizando informações de outros indivíduos, gera para ele classificações mais precisas que ele próprio faz.

Surge então a seguinte questão: "***E se as componentes de especificidade e imprecisão de avaliação fossem menores?*** " em outras palavras, se os indivíduos e objetos fossem muito bem caracterizados, de tal forma que a especificidade fosse mínima e a capacidade de avaliação dos indivíduos fosse muito boa? Do ponto de vista numérico: O que acontece se reduzimos sigma de 5 para 2, por exemplo.

O inusitado aqui não é a resposta, ela se encontra em qualquer livro básico de Regressão, o caráter filtrador continua existindo, mas a aplicação deste resultado para esta questão. Ou seja, não importa se o individuo consegue avaliar um objeto, através do seu escore observado, muito próximo do real escore para ele daquele objeto, ou não, o modelo filtra as imperfeições que ele comete e, revela através dos seus erros, como ele classifica. Abaixo apresentamos como ilustração o que ocorre para o individuo 1, quando reduzimos sigma para 2.

Gráfico 8: Classificação dos objetos para o individuo 1 segundo as suas observações e a estimativa do modelo, considerando o desvio padrão = 2 para a soma da variabilidade da especificidade e imprecisão



O MODELO CLÁSSICO DE REGRESSÃO

Formalmente foi considerado o seguinte no exemplo numérico acima.

Tomamos uma amostra aleatória com reposição de 1000 pares de sujeitos e objetos, do conjunto de 100 sujeitos e 100 objetos apresentados segundo suas características na tabela 8. Isto é, em 1000 situações foram apresentados objetos distintos para indivíduos distintos e registrado o escore ($\boldsymbol{y}$).

Lembremos que a função escore é dada por

$$\mathscr{Y} = \mathscr{S}\left[\underset{\sim}{\boldsymbol{w}}(\boldsymbol{s}),\ \underset{\sim}{\boldsymbol{x}}(\boldsymbol{o}) \right] + \varepsilon = \alpha_o+\alpha_1 x_1+\alpha_2 x_2+\alpha_{12} x_1 x_2+\alpha_{11} x_1^2+\alpha_{22} x_2^2 + \varepsilon$$

com $\alpha_i = \beta_{i,0}+ \beta_{i,1} w_1+ \beta_{i,2} w_2$, para todo i , o que implica que a função escore é

$$\mathscr{Y} = \mathscr{S}\left[\underset{\sim}{\boldsymbol{w}}(\boldsymbol{s}),\ \underset{\sim}{\boldsymbol{x}}(\boldsymbol{o}) \right] + \varepsilon = \beta_{0,0}+ \beta_{0,1} w_1+\beta_{0,2} w_2+\beta_{1,0} x_1+\beta_{1,1} w_1 x_1+\beta_{1,2} w_2 x_2 +. \ . \ . +\beta_{22,0}+$$

$\beta_{22,1} w_1 x_2^2+ \beta_{22,2} w_2 x_2^2 + \varepsilon$. Assim temos,

$\mathscr{Y}$ = escore observado

$$\mathbb{Z} = \begin{bmatrix} \underset{\sim}{1}; \underset{\sim}{w_1}; \underset{\sim}{w_2}; \underset{\sim}{x_1}; \underset{\sim}{x_1 w_1}; \underset{\sim}{x_1 w_2}; \underset{\sim}{x_2}; \underset{\sim}{x_2 w_1}; \underset{\sim}{x_2 w_2}; \underset{\sim}{x_1^2}; \\ \underset{\sim}{x_1^2 w_1}; \underset{\sim}{x_1^2 w_2}; \underset{\sim}{x_1 x_2}; \underset{\sim}{x_1 x_2 w_1}; \underset{\sim}{x_1 x_2 w_2}; \underset{\sim}{x_2^2}; \underset{\sim}{x_2^2 w_1}; \underset{\sim}{x_2^2 w_2} \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_{0,0}; \beta_{1,0}; \beta_{2,0}; \beta_{0,1}; \beta_{1,1}; \beta_{2,1}; \beta_{0,2}; \beta_{1,2}; \beta_{2,2}; \beta_{0,11}; \\ \beta_{1,11}; \beta_{2,11}; \beta_{0,12}; \beta_{1,12}; \beta_{2,12}; \beta_{0,22}; \beta_{1,22}; \beta_{2,22} \end{bmatrix},$$

Com as $\boldsymbol{n}$ independentes observações de $\mathscr{Y}$ e os valores associados de $\boldsymbol{z}$ o modelo torna-se em notação matricial:

$$\underset{(1000\ x\ 1)}{Y} = \underset{(1000\ x\ 18)}{Z}\ \underset{(18\ x\ 1)}{\beta} + \underset{(1000\ x\ 1)}{\varepsilon}$$ com as seguintes especificações:

1. $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ e
2. $Cov(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$

ESTIMATIVA POR MÍNIMOS QUADRADOS

Então a estimativa de $\boldsymbol{\beta}$ por mínimos quadrados é dado por (Mais detalhes ver [27])

$$\hat{\beta}_{coluna} = (Z'Z)^{-1} Z' y$$

Sabendo que sobre as considerações gerais do modelo de regressão linear temos $E(\hat{\beta}_{\text{coluna}}) = \boldsymbol{\beta_{coluna}}$, $Cov(\hat{\beta}_{\text{coluna}}) = \sigma^2(\mathbf{Z'Z})^{-1}$ e que Cov ($\hat{\varepsilon}$) = σ^2[**I**- **H**] e também que $E(\hat{\varepsilon}'\ \hat{\varepsilon})$ = (n - 18 - 1) σ^2 então quando ε tem distribuição normal $\hat{\sigma} = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n}$ é o estimador de máxima verossimilhança de σ^2. Além disso, n $\hat{\sigma}^2$ tem distribuição $\sigma^2\ \chi^2_{n-19}$. Permitindo assim a construção de intervalo de confiança para $\hat{\beta}_{\text{coluna}}$ e $\hat{\sigma}^2$. Estimado o vetor $\boldsymbol{\beta_{coluna}}$, transformamos novamente na Matriz $\hat{\beta}$ com dimensão 6 por 3.

Temos então a Matriz **Y** = ($\hat{\beta}$ W)′ X

onde X é a matriz de características dos 100 objetos, com dimensão 3 por 100 e
W é a matriz de características dos 100 sujeitos envolvidos, com dimensão 6 por 100.

Desta forma, a dimensão de **Y** é 100 por 100, onde nas linhas da matriz temos os escores dos objetos para um mesmo individuo e nas colunas da matriz, o escore de um mesmo objeto para os diferentes indivíduos. Assim **Y** permite construir a matriz **R** de classificação (ranking), onde nas linhas desta matriz teremos a classificação de cada objeto para aquele individuo e nas colunas as classificações de um mesmo objeto para todos os individuos.

Assim através da Matriz **R** temos a classificação dos objetos por individuo e também podemos analisar a variabilidade de classificação de um objeto segundo as prioridades dos vários sujeitos. Desta forma, podemos verificar quais objetos que estão "habitando" os primeiros (ou os últimos, dependendo do interesse) lugares segundo o grupo de sujeitos para os quais estamos fazendo a classificação.

CASO GERAL

Vejamos a situação geral com ***m*** objetos e ***n*** sujeitos, considerando ***v*** características dos objetos e ***u*** características dos sujeitos. A função escore geral com termo de segunda ordem para as características dos objetos e termos de primeira ordem para as características dos sujeitos é dada por

$$\mathscr{S}\left[\ \underset{\sim}{\boldsymbol{w}}(\boldsymbol{s}),\ \underset{\sim}{\boldsymbol{x}}(\boldsymbol{o})\ \right] = \alpha_0 + \sum_{i=1}^{v}\alpha_i x_i + \sum_{i>j}^{p}\alpha_{ij}x_i x_j + \sum_{i=1}^{v}\alpha_{i^2}x_i^2$$

$$\text{onde } p = C_{v,2} = \frac{v(v-1)(v-2)!}{2!(v-2)!} = \frac{v(v-1)}{2}\ ,$$

$\alpha_{ki} = \beta_{0,k} + \beta_{1,k}w_{1,k,i} + \beta_{2,k}w_{2,k,i} + ... + \beta_{u,k}w_{u,k,i} \quad \forall k = 1,2,...,v$ e os escores observados são dados por $\mathscr{Y} = \mathscr{S} + \varepsilon$ onde $\varepsilon \sim N(0\ ;\ \sigma^2)$. A função escore geral possui $(u+1)\frac{1}{2}[v^2+3v+2]+1$ parâmetros a serem estimados, assim o tamanho da amostra, necessário para um experimento, deve ter no mínimo esta quantidade de observações.

Seja uma amostra de trios (***s***, ***o***, ***y***), isto é, um objeto, o sujeito que o avalia e o resultado da avaliação (escore atribuído). Estimamos por mínimos quadrados o Vetor de Betas, que terá dimensão $(u+1)\frac{1}{2}[v^2+3v+2]$ e σ. O modelo que estamos trabalhando envolve:

$\mathscr{Y}$ = escore observado

$$\mathbb{Z} = \begin{bmatrix} \underset{\sim}{1};\underset{\sim}{w_1};\underset{\sim}{w_2};...;\underset{\sim}{w_u};\underset{\sim}{x_1};\underset{\sim}{x_1 w_1};...;\underset{\sim}{x_1 w_u};...;\underset{\sim}{x_v};\underset{\sim}{x_v w_1};...;\underset{\sim}{x_v w_u};\underset{\sim}{x_1^2};\underset{\sim}{x_1^2 w_1};...;\underset{\sim}{x_1^2 w_u}; \\ ;...;\underset{\sim}{x_v^2};\underset{\sim}{x_{v1}^2 w_1};...;\underset{\sim}{x_{v1}^2 w_u};\underset{\sim}{x_1 x_2};\underset{\sim}{x_1 x_2 w_1};...;\underset{\sim}{x_1 x_2 w_u};...;\underset{\sim}{x_{v(v-1)}};\underset{\sim}{x_{v(v-1)} w_1};...;\underset{\sim}{x_{v(v-1)} w_u} \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_{0,0};\beta_{1,0};...;\beta_{u,0};\beta_{0,1};\beta_{1,1};...;\beta_{u,1};...;\beta_{0,v};\beta_{1,v};...;\beta_{u,v};\beta_{0,11};\beta_{1,11};...;\beta_{u,11}; \\ \beta_{0,12};\beta_{1,12};...;\beta_{u,12};...;\beta_{0,v(v-1)};\beta_{1,v(v-1)};...;\beta_{u,v(v-1)};...;\beta_{0,v^2};\beta_{1,v^2};...;\beta_{u,v^2} \end{bmatrix},$$

Com as ***n*** independentes observações de $\mathscr{Y}$ e os valores associados de z o modelo torna-se em notação matricial:

$$\underset{(1000\,x\,1)}{Y} = \underset{(1000\,x\,18)}{Z}\,\underset{(18\,x\,1)}{\beta} + \underset{(1000\,x\,1)}{\varepsilon}$$ com as seguintes especificações:

1. $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ e
2. $Cov(\boldsymbol{\varepsilon}) = \sigma^2 \mathtt{I}$

A estimativa, no caso geral, segue a mesma descrição matricial dada na seção anterior. Assim, temos após a estimação por mínimos quadrados a expressão matricial da Matriz de ***Betas*** que define a função escore geral e permite calcular as funções escore específicas para cada sujeito.

Temos desta forma agora uma matriz ***Y***, com dimensão ***n*** por ***m***, onde nas linhas temos a classificação de todo os objetos para cada um dos sujeitos e nas colunas a classificação de um objeto para todos os sujeitos.

CAPÍTULO III -ESTIMAÇÃO PELA FUNÇÃO DE VEROSSIMILHANÇA

INTRODUÇÃO

No capítulo anterior tratamos de situações em que, confrontado com um objeto qualquer $o \in O$, qualquer sujeito $s \in S$ pode fornecer uma avaliação quantitativa – objetiva, embora contaminada de um certo nível de imprecisão – de seu escore absoluto para aquele objeto. Neste contexto, procedimentos usuais de Regressão Linear – com ajuste de modelo por mínimos quadrados – podem ser aplicados.

Neste capítulo trataremos de situações muito mais comuns, em que a produção de estimativas absolutas do escore é impraticável, podendo os sujeitos apenas ordenar entre si – com alguma qualidade – um conjunto pequeno de objetos, segundo sua preferência percebida.

Experimentos baseados unicamente na ordenação de pequenos sub conjuntos de O, por elementos de S, não produzem informação sobre o nível absoluto da função escore, assim β_{00} , β_{01} , β_{02} , . . . , β_{0u} (que produzem α_0 na função escore específica) são não estimáveis. Alem disto, um fator de escala da função escore estará confundido com o parâmetro σ.

Replicações completas – com o mesmo subconjunto de objetos sendo reapresentado para o mesmo sujeito – removeriam o confundimento, criando informação sobre σ. Todavia tal procedimento não será adotado neste capítulo, uma vez que nem β_{00} , β_{01} , β_{02} , . . . , β_{0u}, nem σ são relevantes para a ordenação dos objetos.

A estimação dos parâmetros do modelo será feita pelo critério da função de verossimilhança. A exploração da função de verossimilhança poderá ser, convenientemente, restrita a uma casca esférica centrada na origem, $\sum \beta^2 + \sigma^2 = r^2$, uma

vez que *L* é invariante com relação a um fator de escala sobre os parâmetros, conforme veremos adiante. Na verdade, como σ é positivo, a busca será restrita à metade da casca esférica definida por $\sigma>0$. Neste trabalho adotaremos sempre r =1.

A precisão associada aos estimadores será calculada com base na matriz de informação de Fisher. Uma vez que os tamanhos amostrais considerados serão sempre grandes – pelo menos da ordem de 1000 – as propriedades assintóticas da matriz garantirão a elevada qualidade das aproximações.

Embora o tamanho dos subconjuntos de O apresentados a elementos de S possa ser qualquer, é razoável admitir que, operacionalmente, tamanhos menores sejam mais prováveis de serem classificados com qualidade. Trataremos aqui de experimentos em que a cada um de n sujeitos – escolhidos aleatoriamente de S – é apresentado um par de objetos, selecionados aleatoriamente de O, para ordenação.

O EXPERIMENTO

Conforme propostos neste trabalho, os modelos bipolares de regressão envolvem, geralmente, um número muito grande de parâmetros. Como vimos, um modelo envolvendo vetores $\underset{\sim}{w}$ e $\underset{\sim}{x}$ com dimensões dois, implica, contando-se o σ, em 19 parâmetros. Elevando-se as dimensões para 3, o número de parâmetros cresce para 45. Situações típicas, com dimensões 5 e 6 para $\underset{\sim}{w}$ e $\underset{\sim}{x}$, respectivamente, implicarão em nada menos que 421 parâmetros para o modelo polinomial completo, e 253 para o modelo polinomial restrito a interações de ordem no máximo igual a 2 entre as componentes de $\underset{\sim}{x}$. A tabela x dá o número total de parâmetros envolvidos, para uma variedade maior de valores u e v.

Tabela 10: Número total de parâmetros envolvidos pelo modelo, em função de u e v.

Dimensão do sujeito	Dimensão do objeto	Número de parâmetros envolvidos	
		Modelo completo	Modelo restrito
2	2	19	19
3	4	81	61
u	v	$(u+1)(2^v+v)+1$	$(u+1)\frac{1}{2}(v^2+3v+2)+1$

Assim, o problema da estimativa de modelos bipolares de classificação por máxima verossimilhança é, necessariamente, um problema computacionalmente intensivo, e demanda amostras grandes. Não deverão ser considerados tamanhos amostrais inferiores a algumas centenas. Consideramos estudos envolvendo pelo menos 1000 unidades amostrais, como os contextos mínimos para a aplicação confortável das idéias aqui propostas.

Os parâmetros do grupo β podem ser, de forma muito conveniente e didática, dada a sua estrutura em blocos, serem representados na forma matricial:

$$\mathrm{B}' = \begin{bmatrix} \beta_{1,0} & \beta_{2,0} & \cdots & \beta_{vv,0} \\ \beta_{1,1} & \beta_{2,1} & \cdots & \beta_{vv,1} \\ \vdots & \vdots & \cdots & \vdots \\ \beta_{1,u} & \beta_{2,u} & \cdots & \beta_{vv,u} \end{bmatrix}$$

contudo, para maior consistência com o desenvolvimento algébrico a seguir, usaremos também a representação vetorial, correspondente ao empilhamento das colunas da matriz B.

Para maior simplicidade da notação nas seções a seguir, trataremos os parâmetros do grupo β e o σ de forma por linha em um único vetor, que denotaremos $\underset{\sim}{\theta}$. Sempre que for necessário, para o completo entendimento do material apresentado, nos referiremos explicitamente a uma ou outra componente deste vetor.

O experimento básico adotado neste capítulo consiste da amostragem de n sujeitos $s \in S$ e, para cada sujeito amostrado, uma amostra de dois objetos $o \in O$. Assim, cada unidade amostrada comporá um trio formado por um sujeito s e dois objetos, o_1 e o_2. Cada sujeito deverá ordenar, segundo suas preferências, os seus dois objetos associados, a partir do conhecimento dos respectivos perfis, $\underset{\sim}{x}_1$ e $\underset{\sim}{x}_2$. Os dados experimentais podem então ser ordenados na forma tabular abaixo.

Tabela 11: Apresentação dos dados experimentais

Trio	Características do Sujeito [$\underset{\sim}{w}$]		Características do Objeto [$\underset{\sim}{x}$]		Escolha
	w_1	w_2	x_1	x_2	
1	w_{11}	w_{21}	x_{11}	x_{21}	t_1
2	w_{12}	w_{22}	x_{12}	x_{22}	t_2
3	w_{13}	w_{23}	x_{13}	x_{23}	t_3
n	w_{1n}	w_{2n}	x_{1n}	x_{2n}	t_n

Na tabela acima, $t_i = 1$ se o i-ésimo indivíduo optou pelo primeiro dos objetos que lhe foram apresentados, e $t_i = 2$ no caso contrário.

No processo de ordenação do par de objetos, por parte de um sujeito nós admitiremos o seguinte processo mental implícito:

- ✓ Com base no conhecimento de $\underset{\sim}{x}_1$ e $\underset{\sim}{x}_2$ (características dos objetos 1 e 2, respectivamente), o sujeito avalia mentalmente os escores correspondentes como:

$$y_i = Y_i(\boldsymbol{s}, \boldsymbol{o}) + \varepsilon_i \text{ onde } \varepsilon_i \sim N(0; \sigma^2) \ \forall \ \boldsymbol{s}, \boldsymbol{o}, \text{ para i=1,2}$$

- ✓ O diferencial percebido, d, é igual ao diferencial real D, contaminado pelos dois erros, ε_1 e ε_2, que assumimos independentes. Assim,

$$d = y_2 - y_1 = Y_2(\boldsymbol{s}, \boldsymbol{o}) + \varepsilon_2 \quad Y_1(\boldsymbol{s}, \boldsymbol{o}) - \varepsilon_1 =$$

$$= Y_2 - Y_1 + \varepsilon_2 - \varepsilon_1 =$$

$$= D + (\varepsilon_2 - \varepsilon_1) =$$

$$= D + \varepsilon$$

onde $\varepsilon \sim N(0 ; 2\sigma^2)$ e

$$D = Y_2(s, o) - Y_1(s, o) =$$

$$= a_1(x_{1,2} - x_{1,1}) + a_2(x_{2,2} - x_{2,1}) + \ldots + a_{vv}(x_{v,2}^2 - x_{v,1}^2)$$

é o diferencial verdadeiro entre os dois objetos, conforme os valores e preferências do sujeito s. Neste contexto de ordenação de dois objetos, D é o sinal e ε o ruído.

Vamos agora definir δ, a forma padronizada de *d*, dividindo-o por seu desvio padrão, $\sqrt{2}\,\sigma$:

$$\delta = \frac{d}{\sqrt{2}\sigma} = \frac{D+\varepsilon}{\sqrt{2}\sigma} = \frac{D}{\sqrt{2}\sigma} + z \text{ onde } z \sim N(0, 1)$$

$$\delta = \frac{1}{\sqrt{2}\sigma}\{a_1(x_{1,2} - x_{1,1}) + a_2(x_{2,2} - x_{2,1}) + \ldots + a_{vv}(x_{v,2}^2 - x_{v,1}^2\} + z =$$

$$= \frac{1}{\sqrt{2}\sigma}\{a_1(x_{1_{máximo}} - x_{1_{mínimo}})\frac{(x_{1,2} - x_{1,1})}{(x_{1_{máximo}} - x_{1_{mínimo}})} + \ldots + a_{vv}(x_{vv_{máximo}} - x_{vv_{mínimo}})\frac{(x_{v,2}^2 - x_{v,1}^2)}{(x_{vv_{máximo}} - x_{vv_{mínimo}})}\} + z$$

Assim, δ é a distância padronizada entre os dois objetos, o_1 e o_2, aos olhos do sujeito s. Definindo

$$\gamma_i = \frac{a_i}{\sqrt{2}\sigma}(x_{i_{máximo}} - x_{i_{mínimo}}) = \frac{\beta_{0,i} + \beta_{1,i}w_1 + \beta_{2,i}w_2 + \cdots + \beta_{u,i}w_u}{\sqrt{2}\sigma}(x_{i_{máximo}} - x_{i_{mínimo}})$$

temos

$$\delta = \gamma_1 \frac{(x_{1,2} - x_{1,1})}{(x_{1_{máximo}} - x_{1_{mínimo}})} + \ldots + \gamma_{vv} \frac{(x_{v,2}^2 - x_{v,1}^2)}{(x_{v_{máximo}}^2 - x_{v_{mínimo}}^2)} + z$$

Finalmente, definindo

$$\pi_i = \frac{x_{i,2} - x_{i,1}}{x_{i_{máximo}} - x_{i_{mínimo}}}$$

chegamos a

$$\delta = \gamma_1\pi_1 + \gamma_2\pi_2 + ... + \gamma_{i,j}\pi_{i,j} + ... + \gamma_{vv}\pi_{vv} + z$$

Definindo

$$\xi = \xi\left(\underset{\sim}{\theta}\right) = \gamma_1\pi_1 + \gamma_2\pi_2 + ... + \gamma_{i,j}\pi_{i,j} + ... + \gamma_{vv}\pi_{vv}$$

chega-se a

$$\delta = \xi + z \text{ onde } \delta \sim N(\xi, 1)$$

Desta forma, sempre que for conveniente, poderemos nos referir aos vetores $\underset{\sim}{\gamma}$ e $\underset{\sim}{\pi}$, equidimensionais, e ao vetor $\underset{\sim}{\xi}$, de dimensão n, que incorporam os respectivos componentes escalares.

Para a determinação da expressão da função de verossimilhança, devemos calcular – dados um sujeito s e dois objetos, o_1 e o_2 – a probabilidade do sujeito dar uma certa ordenação t aos dois objetos, como função das características do trio (s, o_1, o_2): $\underset{\sim}{w}$, $\underset{\sim}{x_1}$ e $\underset{\sim}{x_2}$.

$$P(t/\underset{\sim}{w}, \underset{\sim}{x_1}, \underset{\sim}{x_2}) = P(t/\xi) = P(t=1/\xi)\, I_{t=1} + P(t=2/\xi)\, I_{t=2} =$$

$$= \Phi(-\xi)\, I_{t=1} + \Phi(\xi)\, I_{t=2} =$$

$$= \Phi[(-1)^t \xi]$$

onde $I_{t=i}$ é a função indicadora de t=i e $\Phi(\,.\,)$ é a função de distribuição acumulada de probabilidades da normal padrão.

A função de verossimilhança para uma observação pode então ser escrita como

$$L_1(\xi/t) = \Phi[(-1)^t \xi]$$

logo, para n observações independentes,

$$L_n\left(\underset{\sim}{\xi}/\underset{\sim}{t}\right) = \prod_{i=1}^{n} \Phi[(-1)^{t_i} \xi_i]$$

A função log-verossimilhança, o logaritmo natural de L_n é então dada por:

$$\ell_n\left(\underset{\sim}{\xi}/\underset{\sim}{t}\right) = \ln(L_n(\underset{\sim}{\xi}/\underset{\sim}{t})) =$$

$$= \ln\left\{\prod_{i=1}^{n} \Phi\left[(-1)^{t_i} \xi_i\right]\right\} = \sum_{i=1}^{n} \ln\left\{\Phi\left[(-1)^{t_i} \xi_i\right]\right\}$$

Assim, definimos como o estimador de máxima verossimilhança de $\underset{\sim}{\theta}$ ao vetor $\hat{\underset{\sim}{\theta}}$ que maximiza a função de log-verossimilhança $\ell_n\left(\xi\left(\underset{\sim}{\theta}\right)/\underset{\sim}{t}\right)$.

O estimador de máxima verossimilhança $\hat{\underset{\sim}{\theta}}$ não admite uma expressão algébrica fechada, devendo ser avaliado através de processo numérico de maximização de $\ell_n\left(\xi\left(\underset{\sim}{\theta}\right)/\underset{\sim}{t}\right)$ em $\underset{\sim}{\theta}$. Para a matriz de covariância de $\hat{\underset{\sim}{\theta}}$, empregaremos os resultados assintóticos de Fisher [9].

MATRIZ DE INFORMAÇÃO DE FISHER

Nesta seção vamos construir as expressões algébricas para os termos da Matriz de Informação de Fisher para $\hat{\underset{\sim}{\theta}}$, e explorar suas propriedades mais interessantes. Partimos de algumas definições preliminares:

$$\Delta_i = x_{i\max} - x_{i\min} \quad \forall i = 1,2,...,v \quad (1)$$

$$\Delta_{k_1,k_2} = x_{k_1\max}x_{k_2\max} - x_{k_1\min}x_{k_2\min} \quad (2)$$

$$\pi_{k,i} = \frac{x_{k,2,i} - x_{k,1,i}}{\Delta_k} \quad \forall k = 1,2,...,v \quad (3)$$

$$\pi_{k_1,k_2,i} = \frac{x_{k_1,2,i}x_{k_2,2,i} - x_{k_1,1,i}x_{k_2,1,i}}{\Delta_{k_1,k_2}} \quad (4)$$

$$\gamma_{k,i} = \frac{a_{k,i}\Delta_k}{\sqrt{2}\sigma} \quad (5)$$

$$\gamma_{k_1,k_2,i} = \frac{a_{k_1,k_2,i}\Delta_{k_1,k_2,i}}{\sqrt{2}\sigma} \quad (6)$$

$$a_{ki} = \beta_{0,k} + \beta_{1,k}w_{1,k,i} + \beta_{2,k}w_{2,k,i} + ... + \beta_{u,k}w_{u,k,i} \quad \forall k = 1,2,...,v \quad (7)$$

$$\xi_i = \gamma_{1,i}\pi_{1,i} + \gamma_{2,i}\pi_{2,i} + ... + \gamma_{v,i}\pi_{v,i} + \gamma_{11,i}\pi_{11,i} + \gamma_{12,i}\pi_{12,i} ... + \gamma_{vv,i}\pi_{vv,i} = \underset{\sim}{\gamma}' \underset{\sim}{\pi} \quad (8)$$

Assim, para uma observação, a função de log-verossimilhança é

$$\ell_1(\underset{\sim}{\theta}/\underset{\sim}{t}) = \ell_1(\underset{\sim}{\beta}, \sigma / \underset{\sim}{t}) = \log(L_1(\underset{\sim}{\theta}/\underset{\sim}{t})) = \log\{\Phi[(-1)^t \xi]\}$$

e a Matriz de Informação de Fisher para uma observação é

$$\mathrm{I}_1 = E\,[-\frac{\partial^2}{\partial \underset{\sim}{\theta}^2} \ell_1(\underset{\sim}{\theta} / \underset{\sim}{t})\,]$$

Então vamos calcular a primeira derivada da função de log-verossimilhança, em relação ao vetor de parâmetros $\underset{\sim}{\theta}$.

$$[\frac{\partial}{\partial \theta}\,\ell_1] = \frac{\partial}{\partial \xi}\,\ell_1 \frac{\partial \xi}{\partial \beta_{k_1,k_2}} = \frac{1}{\sqrt{2}\sigma}(-1)^t \;\frac{\varphi((-1)^t \xi)}{\Phi((-1)^t \xi)} \frac{\partial \xi}{\partial \beta_{k_1,k_2}}$$

onde k_1=1,2,…, u e k_2=0,1,2,…, vv

Por exemplo, $\frac{\partial \xi}{\partial \beta_{1,1}} = w_1 d_{x_1}$, $\frac{\partial \xi}{\partial \beta_{1,2}} = w_1 d_{x_1 x_2}$ ou ainda o termo geral, $\frac{\partial \xi}{\partial \beta_{i,j}} = w_i d_{x_j}$.

Assim, o termo geral da primeira derivada da função de log-verossimilhança é

$$[\frac{\partial}{\partial \theta}\,\ell_1] = \frac{\partial}{\partial \xi}\,\ell_1 \frac{\partial \xi}{\partial \beta_{k_1,k_2}} = \frac{1}{\sqrt{2}\sigma}(-1)^t \;\frac{\varphi((-1)^t \xi)}{\Phi((-1)^t \xi)} w_{k_1} d_{x_{k_2}}$$

Portanto, a primeira derivada da função de log-verossimilhança em termos de ξ, a menos do fator de escala $\frac{1}{\sqrt{2}\sigma}$ é

$$\frac{\partial}{\partial \xi}\mathrm{l}_1 = (-1)^t \;\frac{\varphi((-1)^t \xi)}{\Phi((-1)^t \xi)} = \begin{cases} -\dfrac{\varphi(\xi)}{1-\Phi(\xi)} = -H(\xi) & para\ t = 1, \\ \dfrac{\varphi(-\xi)}{1-\Phi(-\xi)} = H(-\xi) & para\ t = 2, \end{cases}$$

Na expressão acima, H é a função de risco (*hazard rate function*) para a normal padrão.

Veja que, conforme esperado,

$$\mathrm{E}\left(\frac{\partial}{\partial\xi}\ell_1\right) \cong -H(\xi)\,\Phi(-\xi) + H(-\xi)\,\Phi(\xi)) = -\frac{\varphi(\xi)}{1-\Phi(\xi)}\,\Phi(-\xi) + \frac{\varphi(-\xi)}{1-\Phi(-\xi)}\,\Phi(\xi)$$

$$= -\frac{\varphi(\xi)}{\Phi(-\xi)}\,\Phi(-\xi) + \frac{\varphi(-\xi)}{\Phi(\xi)}\,\Phi(\xi) = -\varphi(\xi) + \varphi(-\xi) = = -\varphi(\xi) + \varphi(\xi) = 0$$

Vejamos agora o termo geral da segunda derivada de ℓ_1.

$$\frac{\partial}{\partial\beta_{k,l}}\left(\frac{\partial}{\partial\beta_{i,j}}\mathrm{l}_1\right) = \frac{1}{2\sigma^2}\,w_i d_{x_j}\,w_k d_{x_l}\begin{cases}\dfrac{\partial}{\partial\xi}(-H(\xi)) & t=1\\[2ex] \dfrac{\partial}{\partial\xi}(H(-\xi)) & t=2\end{cases}$$

A menos do termo $\frac{1}{2\sigma^2}\,w_i d_{x_j}\,w_k d_{x_l}$ temos que a segunda derivada de ℓ_1 em termos de ξ é

$$\frac{\partial^2}{\partial\xi^2}\ell_1 = \begin{cases}\dfrac{\partial}{\partial\xi}\left[-\dfrac{\varphi(\xi)}{1-\Phi(\xi)}\right] = -\dfrac{\xi\varphi(\xi)[1-\Phi(\xi)]+\varphi(\xi)\varphi(\xi)}{[1-\Phi(\xi)]^2} = -\xi\dfrac{\varphi(\xi)}{1-\Phi(\xi)} - \left[\dfrac{\varphi(\xi)}{1-\Phi(\xi)}\right]^2\\[2ex] \dfrac{\partial}{\partial\xi}\left[\dfrac{\varphi(-\xi)}{1-\Phi(-\xi)}\right] = \dfrac{-\xi\varphi(-\xi)[1-\Phi(\xi)]+\varphi(-\xi)\varphi(-\xi)}{[1-\Phi(-\xi)]^2} = -\xi\dfrac{\varphi(-\xi)}{1-\Phi(-\xi)} + \left[\dfrac{\varphi(-\xi)}{1-\Phi(-\xi)}\right]^2\end{cases}$$

$$\frac{\partial^2}{\partial\xi^2}\mathrm{l}_1 = \begin{cases} -\xi\,H(\xi) - H^2(\xi) = -H(\xi)(\xi + H(\xi)) & para\ t=1\\[2ex] \xi\,H(-\xi) - H^2(-\xi) = H(-\xi)(\xi - H(-\xi)) & para\ t=2\end{cases}$$

Assim temos, a menos da constante $\frac{1}{2\sigma^2}\,w_i d_{x_j}\,w_k d_{x_l}$, a esperança da segunda derivada de ℓ_1.

$$\mathrm{E}\left(-\frac{\partial^2}{\partial\xi^2}\ell\right) = (-H(\xi)(\xi + H(\xi)))\ \mathrm{P}(t=1) + (H(-\xi)(\xi - H(-\xi)))\ \mathrm{P}(t=2) =$$

$$= -H(\xi)[\xi + H(\xi)]\ \Phi(-\xi) + H(-\xi)[\xi - H(-\xi)]\ \Phi(\xi) =$$

$$= \varphi(\xi)(\xi + H(\xi)) + \varphi(\xi)[\xi - H(-\xi)] = \varphi(\xi)[H(\xi) + H(-\xi)]$$

Portanto, o termo geral, (i, j), da Matriz de Informação de Fisher, para uma observação, é

$$I_{1_{i,j}} = E\{\frac{\partial}{\partial\theta}\ell_1(\underset{\sim}{\theta}/\underset{\sim}{t})\}^2 = -E\{\frac{\partial^2}{\partial\theta^2}\log(L_1(\underset{\sim}{\theta}/\underset{\sim}{t})\} = \frac{1}{\sqrt{2}\sigma} w_i d_{x_i} w_j d_{x_j} \varphi(\xi)\left[H(\xi) + H(-\xi)\right]$$

e, para o caso de n observações independentes,

$$I_{n_{i,j}} = nE\{\frac{\partial}{\partial\theta}\ell_1(\underset{\sim}{\theta}/\underset{\sim}{t})\}^2 = -nE\{\frac{\partial^2}{\partial\theta^2}\log(L_1(\underset{\sim}{\theta}/\underset{\sim}{t})\} = \frac{n}{\sqrt{2}\sigma} w_i d_{x_j} w_k d_{x_l} \varphi(\xi)\left[H(\xi) + H(-\xi)\right]$$

O PROCESSO DE BUSCA

A determinação das estimativas de máxima verossimilhança dos parâmetros implica na busca em um espaço de dimensão igual à do vetor $\underset{\sim}{\theta}$. Nesta busca o objetivo é encontrar o ponto $\underset{\sim}{\hat{\theta}}$ do espaço dos $\boldsymbol{\theta}$, no qual a função de log-vesossimilhança atinge o valor máximo. Algumas considerações sobre o comportamento da função de log-verossimilhança neste caso, no espaço considerado, permite uma substancial simplificação do problema.

Como o modelo considerado trabalha não com escores individuais, mas com diferenças padronizadas de escores, a função de log-verossimilhança é invariante com relação a fatores de escala. Assim,

$$\ell_n\left(\xi\left(\underset{\sim}{\theta}\right)/\underset{\sim}{t}\right) = \ell_n\left(\xi\left(k\,\underset{\sim}{\theta}\right)/\underset{\sim}{t}\right)$$

para qualquer k real diferente de zero. Podemos, pois, restringir a busca a uma casca esférica, centrada na origem, de um raio pré determinado.

O volume de uma esfera de raio r e dimensão ν [8] é dada por

$$V_\nu(r) = \frac{\pi^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}+1\right)} r^\nu$$

Para um raio fixo – unitário, por exemplo – este volume converge para zero quando ν cresce para ∞. Assim, o volume da hiper esfera de dimensão ν é crescentemente insignificante frente ao volume do hiper cubo circunscrito, à medida em que cresce a dimensão como ilustra a tabela abaixo. Nela são apresentados, para diversos valores de ν, a razão entre $V_\nu(r)$ e o volume do hipercubo circunscrito, $(2r)^\nu$, que denominaremos ρ.

Tabela 12: Distribuição de alguns valores da razão entre o volume do hipercubo circunscrito a esfera em função do número de parâmetros

ν	1	2	3	4	5	10	15	20	25
ρ	1	0.786	0.5236	0.3084	0.1645	0.00249	$1.2\text{x}10^{-5}$	$2.5\text{x}\ 10^{-8}$	$2.8\text{x}10^{-11}$

Para ν = 100, uma dimensão relativamente modesta no contexto de modelos bipolares como proposto, o volume da hiper esfera de raio unitário é igual a $2.368\text{x}10^{-40}$, e a relação ρ é igual a $1.868\text{x}10^{-70}$.

Uma abordagem aparentemente eficaz para a exploração da função de log-verossimilhança na superfície (hiper) esférica considerada, começaria salpicando - aleatoriamente e com distribuição uniforme – pontos sobre a superfície esférica, para a determinação de um bom ponto de partida; e em seguida explorando, sobre a superfície esférica, uma vizinhança arbitrariamente estreita deste ponto. A escolha aleatória de pontos uniformemente distribuídos sobre a superfície esférica se daria através de uma operação em três passos computacionalmente triviais:

[8] APOSTOL, TOM M. **Calculus vol. II** 2º ed. John Wiley & Sons New York 1969 pag. 411

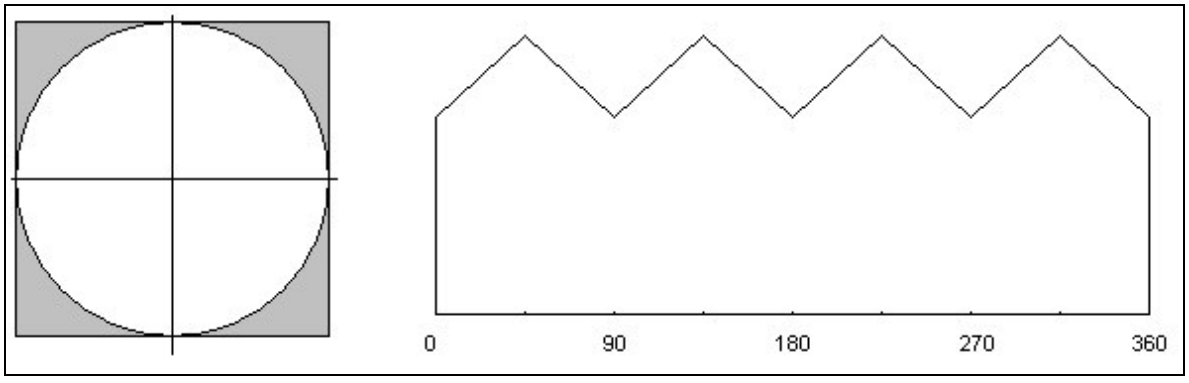
1. Seleção, ao acaso, de um ponto no cubo circunscrito;
2. Rebatimento – projeção radial – deste ponto para a superfície esférica;
3. Calculo da função de log verossimilhança para o ponto resultante.

A repetição desta operação algumas milhares de vezes – uma tarefa menor para um bom micro computador – nos forneceria, naquela com maior valor para log verossimilhança, um bom ponto inicial para exploração da superfície esférica.

Esta abordagem para a definição de um ponto de partida, contudo, tem inconvenientes que a desqualificam. Para começo, ela não distribui pontos uniformemente sobre a superfície esférica.

Em dimensões elevadas, como vimos, o volume da esfera é insignificante em relação ao do cubo circunscrito, assim, a vasta maioria dos pontos gerados no primeiro passo, se situarão fora da esfera. Os pontos fora da esfera, ao se rebaterem sobre a superfície desta, se concentrarão mais nos pontos correspondentes aos vértices, e menos nos pontos de tangenciamento esfera-cubo. A figura abaixo ilustra este problema, para o caso de dimensão dois.

Figura 1: Comportamento da concentração de pontos fora da esfera em função do ângulo



Este problema se agrava muito com o aumento de ν, ao ponto de, em dimensões elevadas, praticamente só se gerar pontos numa vizinhança estreita das projeções dos vértices, com as vizinhanças dos pontos de tangência sendo raramente visitados.

Por outro lado, a exclusão *a priori*, no Passo 1, dos pontos exteriores à esfera – os responsáveis pela distorção na distribuição de pontos sobre a superfície esférica – seria enormemente ineficiente, dada a insignificância da esfera frente ao cubo circunscrito.

A abordagem do problema por coordenadas polares, além de destruir a simplicidade e elegância do processo, traria sérias complicações de cálculo que também lhe comprometeria drasticamente a eficiência. A geração, numericamente eficiente, de pontos uniformemente distribuídos sobre uma superfície esférica de dimensão elevada se revelou, assim, um problema não trivial.

Consideramos aqui uma abordagem alternativa: Percorrer os vértices do cubo – também uma operação trivial, embora proibitivamente extensa, mesmo para dimensões não muito elevadas – e adotar como ponto de partida a projeção sobre a superfície esférica do vértice de maior valor da função de log verossimilhança. Um hipercubo de dimensão 15, como o envolvido no exemplo neste capítulo, possui 2^{15} – ou 32768 – vértices; a exploração de todos eles é factível e, num micro computador com processador de 2.5GHz, tomou menos que 5 minutos, e o esforço de programação do algoritmo é mínimo.

Esta abordagem foi usada no Exemplo Numérico, deste capítulo, para a determinação do ponto inicial da busca.

Definido o ponto inicial, passa-se ao estágio seguinte, onde pontos são gerados uniformemente sobre uma calota esférica centrada no ponto de partida. O ângulo da calota é arbitrário, e controlado pelo módulo do vetor diferença entre o novo ponto gerado e a ponto de partida, ou central. O vetor diferença é tangente à esfera, no ponto de partida, com a direção no plano tangente variando uniformemente de 0 a 2π, e módulo distribuído segundo uma distribuição triangular entre 0 e k. O módulo k, define o ângulo de abertura da

calota esférica, ou da vizinhança sendo explorada; por exemplo, k=1 define uma calota de meia abertura igual a 45^o.

Vejamos a seguir um exemplo numérico desta situação, através de uma simulação

UM EXEMPLO NUMÉRICO

Nesta simulação iremos gerar trios formados por dois objetos – sorteados aleatoriamente do conjunto de objetos a serem ordenados – e um sujeito, sorteado aleatoriamente do conjunto de sujeitos para o qual destina-se a ordenação. A escolha do sujeito: objeto 1 ou objeto 2, registrada.

Com uma amostra de n=1000 trios, teremos 1000 sujeitos e 2000 objetos, compondo 1000 pares, definidos por duas características (x_1 , x_2) com valores entre 0 e 1, gerados uniformemente. Cada par de objetos será classificado por um grupo de 1000 sujeitos, especificados por suas características (w_1 , w_2), também com valores entre 0 e 1.

Seja B , a função escore geral real dada, por exemplo, pela seguinte matriz:

$$\mathrm{B} = \begin{bmatrix} 30.0 & 10.0 & -5.0 \\ 22.0 & -2.5 & 1.8 \\ 14.0 & 1.5 & 0.8 \\ 6.0 & -0.2 & -0.4 \\ 3.0 & 0.3 & 0.1 \\ 6.0 & -2.0 & 2.0 \end{bmatrix}$$

Aplicando a matriz **B** ao trio (s, o_1, o_2), com características: $\underset{\sim}{w}$, $\underset{\sim}{x_1}$ e $\underset{\sim}{x_2}$ temos o escore real para cada objeto do par apresentado para cada sujeito. Geramos arbitrariamente um desvio $\varepsilon \sim N(0, 4^2)$, que é resultado da composição da especificidade e imprecisão da obtenção do escore de cada objeto para cada sujeito.

Usando o critério de máxima verossimilhança e o processo de busca anteriormente descrito, encontramos uma estimativa para matriz **B** e utilizando a Matriz de Informação de Fisher temos a variância de cada parâmetro estimado. A seguir apresentamos para este exemplo a matriz estimada e sua correspondente matriz de variância. O importante aqui é a comparação da classificação de um sujeito qualquer dada pela matriz **B** acima, a classificação gerada por ele considerando as componentes de especificidade e imprecisão da função escore e a classificação gerada segundo a matriz estimada, a qual apresentamos a seguir.

Para este exemplo, a matriz **B** estimada, na esfera de raio 1, é:

$$\hat{B} = \begin{bmatrix} 0.014 & 0.139 & 0.075 \\ -0.271 & 0.272 & 0.074 \\ 0.184 & -0.121 & 0.436 \\ -0.216 & -0.234 & -0.235 \\ -0.353 & 0.307 & 0.312 \end{bmatrix}$$

com $\hat{\sigma} = 0.331$

Como a Matriz de Informação é simétrica apresentamos, a seguir, somente a metade superior.

$$I = \begin{bmatrix} 247.4 & 127.3 & 119.4 & 25.2 & 13.1 & 11.4 & 242.9 & 124.5 & 115.9 & 140.3 & 72.9 & 68.6 & 27.1 & 13.7 & 11.5 \\ & 86.3 & 60.9 & 13.1 & 9.4 & 4.3 & 124.5 & 83.9 & 58.6 & 72.9 & 50.1 & 34.6 & 13.7 & 9.6 & 3.9 \\ & & 78.2 & 11.4 & 4.3 & 6.1 & 115.9 & 58.6 & 75.2 & 68.6 & 34.6 & 44.3 & 11.5 & 3.9 & 5.6 \\ & & & 261.2 & 138.5 & 139.6 & 25.4 & 13.9 & 12.3 & 140.2 & 74.5 & 73.0 & 265.4 & 141.8 & 140.7 \\ & & & & 96.0 & 71.8 & 13.9 & 10.1 & 5.0 & 74.5 & 51.8 & 36.7 & 141.8 & 98.5 & 72.7 \\ & & & & & 96.0 & 12.3 & 5.0 & 7.0 & 73.0 & 36.7 & 48.9 & 140.7 & 72.7 & 96.3 \\ & & & & & & 256.5 & 131.2 & 121.4 & 138.6 & 72.1 & 67.2 & 26.3 & 13.9 & 11.7 \\ & & & & & & & 88.0 & 60.9 & 72.1 & 49.4 & 33.6 & 13.9 & 10.1 & 4.3 \\ & & & & & & & & 78.1 & 67.2 & 33.6 & 43.0 & 11.7 & 4.3 & 5.9 \\ & & & & & & & & & 162.6 & 85.4 & 82.2 & 143.6 & 76.7 & 73.5 \\ & & & & & & & & & & 59.0 & 41.7 & 76.7 & 53.5 & 37.0 \\ & & & & & & & & & & & 54.0 & 73.5 & 37.0 & 48.6 \\ & & & & & & & & & & & & 288.1 & 154.5 & 150.8 \\ & & & & & & & & & & & & & 107.5 & 78.2 \\ & & & & & & & & & & & & & & 102.5 \end{bmatrix}$$

Da mesma forma, apresentamos somente a metade superior da matriz de Variância dos Betas. Lembramos que os parâmetros associados a a_0 que correspondia a primeira linha da matriz beta completa, não são estimados pois não interferem na diferença dos escores dos objetos.

$$\hat{\Sigma}_B = \begin{bmatrix} 0.56 & -0.48 & -0.49 & 0.03 & 0.02 & -0.07 & -0.46 & 0.39 & 0.40 & -0.17 & 0.15 & 0.15 & 0.05 & -0.08 & 0.00 \\ & 0.83 & 0.09 & 0.01 & 0.04 & -0.04 & 0.39 & -0.66 & -0.09 & 0.15 & -0.28 & -0.03 & -0.07 & 0.09 & 0.05 \\ & & 0.91 & -0.06 & -0.06 & 0.22 & 0.41 & -0.09 & -0.73 & 0.15 & -0.03 & -0.30 & -0.01 & 0.07 & -0.07 \\ & & & 0.44 & -0.34 & -0.40 & 0.03 & -0.05 & 0.01 & -0.12 & 0.08 & 0.11 & -0.35 & 0.28 & 0.32 \\ & & & & 0.73 & -0.06 & -0.06 & 0.08 & 0.03 & 0.08 & -0.21 & 0.05 & 0.28 & -0.58 & 0.03 \\ & & & & & 0.88 & 0.02 & 0.01 & -0.09 & 0.11 & 0.05 & -0.28 & 0.32 & 0.03 & -0.69 \\ & & & & & & 0.47 & -0.40 & -0.41 & 0.02 & -0.01 & -0.03 & -0.03 & 0.06 & -0.01 \\ & & & & & & & 0.70 & 0.07 & -0.02 & -0.02 & 0.06 & 0.05 & -0.06 & -0.03 \\ & & & & & & & & 0.75 & -0.02 & 0.05 & 0.02 & 0.00 & -0.05 & 0.08 \\ & & & & & & & & & 0.31 & -0.26 & -0.26 & -0.03 & 0.04 & 0.01 \\ & & & & & & & & & & 0.57 & -0.05 & 0.04 & -0.06 & -0.02 \\ & & & & & & & & & & & 0.57 & 0.02 & -0.02 & 0.01 \\ & & & & & & & & & & & & 0.36 & -0.30 & -0.32 \\ & & & & & & & & & & & & & 0.60 & -0.01 \\ & & & & & & & & & & & & & & 0.68 \end{bmatrix}$$

Dada a matriz Beta estimada temos para cada individuo específico sua função escore, dada pela multiplicação da matriz beta pelo vetor de características deste individuo.

Vejamos, por exemplo, a classificação de um individuo, sorteado aleatoriamente do conjunto de indivíduos. Ele tem as características (w_1, w_2) = (0.929303 , 0.137646) e considerando a Matriz Beta original temos a sua função escore real, dada por:

$$Y = 38.6 + 19.9x_1 + 15.5x_2 + 5.8x_1^2 + 3.3x_1x_2 + 4.4x_2^2$$

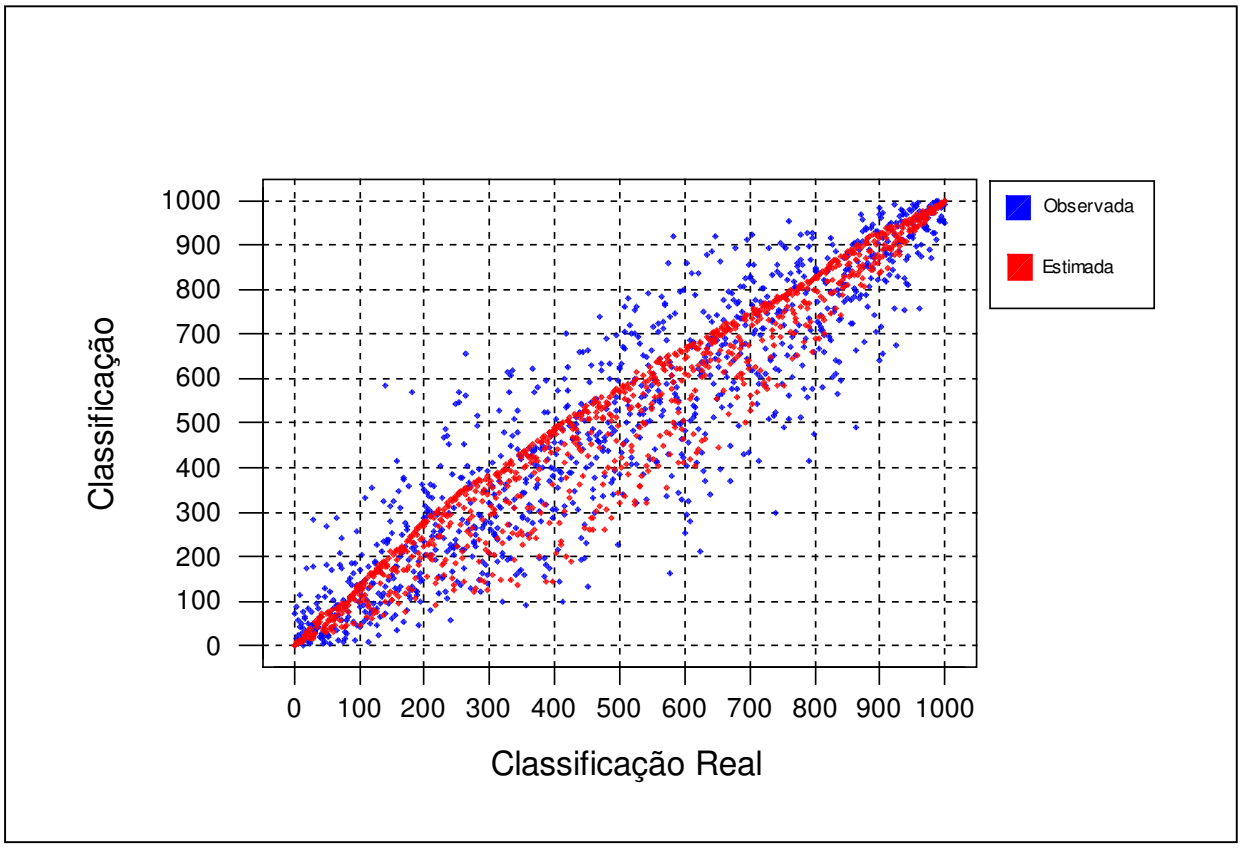
Com a expressão da função escore, dada acima, temos a classificação do conjunto de objetos para o sujeito caracterizado acima.

Com as observações sendo “perturbadas” segundo uma distribuição normal com desvio padrão igual a 4, temos a classificação que seria gerada por este individuo, caso ele tivesse ordenado os objetos segundo o escore percebido por ele.

Dada a matriz beta estimada, temos a função escore estimada para este individuo, que mais importante que a sua expressão é a classificação gerada por ela e comparada com a classificação real e observada.

Apresentamos no gráfico abaixo a classificação observada x a classificação real e a classificação estimada x classificação real, veja que a correlação é maior entre a classificação estimada e a classificação real.

Gráfico 9: Comportamento da classificação real x classificação estimada e classificação real x classificação observada para um sujeito específico



CONCLUSÃO

INTRODUÇÃO

A superioridade da abordagem bipolar para problemas de classificação, sobre o esquema tradicional – que apenas considera as características dos objetos e ignora a variabilidade de características entre os sujeitos – é difícil de ser questionada.

Uma pesquisa na literatura revela indícios de uma insatisfação subjacente com relação ao enfoque tradicional [42] , porem as alternativas, quando propostas, tinham caráter informal e

A inexistência de uma teoria adequada à modelação bipolar, bem como de procedimentos experimentais e de estimação completos impossibilitariam o progresso em uma eventual investida nesta direção.

Neste trabalho, procuramos, primeiro, abrir a discussão sobre a inadequação inerente à abordagem unipolar para problemas de classificação. Esta inadequação ficou sintetizada na pergunta: Para quem – para que grupo específico de cidadãos – estão direcionadas as classificações periódicas que se faz em vários países – Estados Unidos em particular – das principais áreas urbanas, com relação a diversas variantes de Índices de Qualidade de Vida?

A pergunta procede, uma vez que é natural esperar que cidadãos – ou grupos de cidadãos – de características sócio-econômicas, etárias e culturais diferentes possam apresentar perfis de preferências também bastante diferentes. Se, por um lado, Orange County foi classificada como a melhor cidade dos EUA em 1999, com NY ocupando um distante 72^{o}. lugar, por outro, seria facílimo recrutar milhões de cidadãos americanos, que não trocariam Nova York por Orange County por nada deste mundo. Classificação de objetos deve, necessariamente, levar em consideração variáveis dos Sujeitos a que se destina, sob pena de ter um caráter fortemente abstrato.

Neste trabalho buscamos desenvolver uma abordagem metodológica completa, com sólida e rigorosa fundamentação teórica que permitisse a construção de modelos bipolares conceitualmente simples, com estratégias experimentais alternativas, e com procedimentos eficazes para a estimação dos parâmetros associados, a partir dos dados experimentais.

Na questão dos planos experimentais consideramos dois contextos alternativos, um no qual um sujeito consegue fornecer uma avaliação quantitativa absoluta para o seu escore para um objeto, e outra na qual o melhor que ele consegue é ordenar dois objetos que lhes são apresentados. No primeiro caso, o problema da estimativa e inferências sobre os parâmetros do modelo pode se beneficiar da teoria clássica do ajuste de modelos por mínimos quadrados. No segundo caso, desenvolvemos uma abordagem completa através do critério da máxima verossimilhança, e com o emprego da teoria assintótica associada à Matriz de Informação de Fisher, a qual se adequa com perfeição às características especiais do problema.

A determinação das expressões algébricas para os elementos da matriz de informação de Fisher, para o contexto específico dos modelos empregados, se constituiu num desafio considerável em termos dos desenvolvimentos algébricos envolvidos. No Capítulo 3 realizamos as tarefas estruturais básicas daquele desenvolvimento, as quais podem ser estendidas, sem dificuldade a outros contextos correlatos. No final, como sugere a nossa experiência, todo o desenvolvimento algébrico se revelará muito mais tedioso e trabalhoso do que realmente complexo.

Uma característica de início assustadora dos modelos bipolares proposto é o número de parâmetros envolvidos, extraordinariamente alto. Esta característica terá presença certa na maioria dos modelos bipolares duplamente polinomiais, conforme comentamos no Capítulo 3. O elevado número de parâmetros nestes casos coloca um desafio e uma restrição à aplicação destes modelos.

O desafio está nos procedimentos numéricos de busca, que deverão explorar hiper esferas de raios unitários e centradas na origem, em espaços de altas dimensões. O confinamento da busca à casca das hiper esferas, contudo, simplifica bastante o problema de busca, o qual, de qualquer modo, se adequa bem aos recursos computacionais modernos. Neste território, este trabalho deixa sugeridos diversas linhas de pesquisa em rotinas eficientes de buscas nestas cascas hiper esféricas. Neste trabalho mostramos a equivalência desta busca à exploração dos vértices do hiper cubo envoltório e, uma vez localizado o melhor vértice – pelo critério da verossimilhança – a busca subseqüente se restringirá a uma vizinhança do ponto correspondente a este vértice na superfície da hiper esfera. O procedimento é de grande elegância conceitual e operacional, e foi apenas superficialmente explorado neste trabalho, deixando implicitamente sugeridas interessantes linhas de pesquisa subseqüente.

Falamos também de uma limitação importante dos modelos bipolares duplamente polinomiais. O elevado número de parâmetros restringe suas aplicações a contextos onde tamanhos amostrais da ordem de alguns milhares sejam o caso. Estes modelos não são adequados, a não ser em casos muito específicos, a situações onde existam severas restrições ao tamanho amostral. De um modo geral podemos dizer que uma amostra de tamanho mil seria o ponto de partida para a aplicação adequada destes modelos.

Se, por um lado, estas limitações excluem a possibilidade de aplicação dos modelos a diversos contextos onde eles, a menos da questão do tamanho amostral, se ajustariam com perfeição, por outro, existe uma enorme variedade de situações de grande interesse prático, onde grandes massas de dados são a regra.

Neste sentido lembramos as bases de dados do governo federal, onde freqüentemente estas bases têm periodicidade mensal, e consistem de, não raro, milhões de pontos.

Um outro aspecto muito interessante dos modelos bipolares de classificação se refere a uma aparente limitação. Como interpretar os parâmetros de um modelo quando existem, freqüentemente, centenas deles?

Contudo, não perdendo de vista o objetivo central destes modelos, que é o de fornecer funções de critérios de classificação de objetos, que sejam específicas para sujeitos – ou grupos de sujeitos – podemos nos concentrar no valor global da função escore estimada, sem nos preocuparmos com os parâmetros que a constituem. O fato de que estas funções de classificação introduzem uma melhora substancial na qualidade da classificação dos objetos, vis-à-vis uma eventual – e altamente impraticável – alternativa direta – feita pessoalmente pelo sujeito – deve bastar aos propósitos deste trabalho.

ABORDAGEM DO PROBLEMA: TÁTICAS EXPERIMENTAIS E TÉCNICAS DE ESTIMAÇÃO

O ajuste de modelos bipolares de classificação, como nos casos usuais de modelos de regressão, envolvem dois estágios distintos, a coleta de dados e a subseqüente estimativa dos parâmetros com base nos dados experimentais. Existe uma literatura abundante nesta área, referente a delineamentos ótimos, que buscam maximizar a qualidade estatística das estimativas, para um custo experimental – geralmente associado a tamanho amostral – fixo.

No caso de ajuste de modelos por regressão linear o problema de estimação dos parâmetros a partir dos dados, tem solução simples através do critério de mínimos quadrados, que fornece soluções algebricamente fechadas para os estimadores dos parâmetros.

No caso de modelos bipolares duplamente polinomiais, como os explorados nos capítulos 2 e 3, os resultados teóricos existentes para modelos clássicos de regressão linear múltipla podem ser aplicados diretamente, desde que os dados experimentais sejam compatíveis, como tratado no Capítulo 2.

Naquele contexto, quando confrontado com um objeto qualquer do conjunto a ser classificado, qualquer sujeito é capaz de fornecer uma avaliação objetiva, embora imprecisa, de seu escore para aquele objeto. Nestes casos, um delineamento experimental que selecione, por algum critério, n pares (s,o) de um sujeito e um objeto, e anote o escore avaliado do objeto o pelo sujeito s, fornecerá um conjunto de dados (sujeito, objeto, resposta), que se adequa com perfeição ao esquema de regressão linear múltipla, embora com uma estrutura peculiar, que envolve variáveis do sujeito e variáveis do objeto, e um modelo duplamente polinomial imbricado, com dois níveis de parâmetros, denominados α e β.

Esta estrutura imbricada especial abre amplas e interessantes vias de exploração no sentido da construção de delineamentos ótimos, as quais, todavia, por limitação de tempo e por não fazerem parte do escopo pré definido para este projeto, não exploramos.

Situações mais comuns são aquelas em que o máximo que o sujeito consegue é ordenar, pelo critério do escore, um conjunto pequeno de objetos. No Capítulo 3 tratamos destes casos, porem restritos a conjuntos de dois objetos. Um elemento experimental, neste caso, consiste de um sujeito, dois objetos e a ordem escolhida: (s, o_1, o_2, t), onde o s é caracterizado por um vetor $\underset{\sim}{w}$, os objetos o_1 e o_2 pelos vetores $\underset{\sim}{x}_1$ e $\underset{\sim}{x}_2$, e t é igual a 1 ou 2, conforme a escolha de s, entre os dois objetos. Um experimento completo consiste de n desses elementos básicos.

Se, por um lado, esta segunda alternativa experimental é mais geral e de maior aplicabilidade, por outro ela exige tamanhos amostrais maiores, uma vez que a quantidade de informação sobre os parâmetros do modelo, contida em um elemento experimental é menor do que no caso anterior, onde o sujeito fornece uma avaliação absoluta do escore.

A grande generalidade desta abordagem experimental tem ainda um outro custo: o cálculo das estimativas dos parâmetros, a partir dos resultados experimentais, não conta com uma solução algébrica fechada como no caso anterior. A determinação das estimativas envolve a maximização numérica de uma função de log-verossimilhança razoavelmente

complexa, num espaço euclidiano de dimensão elevada. O ajuste de modelos bipolares de classificação é, certamente, um problema computacionalmente intensivo, alem de depender de tamanhos amostrais elevados. Os desafios postos, contudo, são rotineiros, dada a riqueza existente e disponível, tanto de recursos computacionais como em procedimentos gerais para a exploração de superfícies complexas em dimensões elevadas.

De qualquer forma, como no caso dos delineamentos experimentais, os desafios numéricos de localização do máximo de uma função de log-verossimilhança abrem interessantes linhas de pesquisa sobre procedimentos eficientes de exploração de uma superfície esférica de dimensão elevada. Neste trabalho exploramos algumas, mas não nos alongamos muito nestas vias, por não pertencerem ao escopo restrito deste projeto.

CAMPOS DE APLICAÇÃO

Os problemas de classificação são amplamente difundidos na sociedade moderna. A meritocracia implícita em nossos valores sociais, num certo sentido implica na generalização de procedimentos de ordenação de conjuntos – os mais gerais possíveis – de objetos.

Com raras – e freqüentemente toscas – exceções, o problema de classificação que é, normalmente e intrinsecamente, bipolar, é tratado por abordagens exclusivamente unipolares. As razões principais para este estado de coisas são duas:

1. É assim que sempre foi feito, e geralmente não se questiona práticas antigas e bem estabelecidas;
2. A inexistência de suporte teórico e ferramental metodológico apropriados a uma abordagem bipolar.

Neste trabalho procuramos dar uma contribuição no sentido, principalmente, do item 2, acreditando que o item 1 tenderá a vir a reboque, com o tempo e a difusão e expansão destes conceitos e métodos.

Um exemplo de tentativa – singela – de considerar o problema da classificação de objetos com uma abordagem bipolar, são os vestibulares a algumas universidades brasileiras, UNICAMP inclusive. A adoção de pesos diferenciados por faculdade, para as notas das diversas disciplinas[9], se constitui num passo importante para o aperfeiçoamento do processo de seleção. A escolha dos pesos por disciplina, embora apontando qualitativamente em direções provavelmente corretas, foi feita de forma arbitrária, sem a aplicação de critérios quantitativos de otimização das características seletivas do processo de classificação. A direção está provavelmente correta, mas a intensidade pode ser aperfeiçoada.

Acreditamos que a continuação deste estado de coisas não se justifica mais, agora que a discussão foi levantada, e uma proposta objetiva de abordagem do problema da construção de modelos bipolares ótimos foi desenvolvida. A direção dos trabalhos deveria ser no sentido da exploração das bases de dados históricos dos vestibulares, cruzando-os com dados de performance do aluno aprovado ao longo de sua passagem pela universidade e, posteriormente, como membro produtivo da sociedade. O objetivo: a construção de um modelo de classificação que seja ótimo, no sentido de selecionar objetos (os candidatos), melhor ajustados às exigências e necessidades dos sujeitos (os departamentos).

Uma outra linha de aplicação destas estruturas teóricas e instrumentos metodológicos, referem-se à classificação de países, economias, ou comunidades. Esta foi a questão motivacional levantada na introdução deste trabalho e que, de uma forma mais ou menos explícita, permeou todo o desenvolvimento do texto.

[9] Na Unicamp, por exemplo, desde 1995, o vetor de pesos para as disciplinas (1ª.Fase, Português, Matemática, História, Geografia, Biologia, Física, Química, Língua Estrangeira) é igual a (0.16; 0.08; 0.08; 0.08; 0.08; 0.16; 0.08; 0.16; 0.08) para a Faculdade de Medicina enquanto que para a Faculdade de Engenharia Elétrica é igual a (0.16; 0.08; 0.16; 0.08; 0.08; 0.08; 0.16; 0.08; 0.08).

Neste campo, apesar do enfoque jornalístico voltar-se para o sujeito, ele não tem sido considerado no processo de classificação, como mostramos, por exemplo, no Capítulo 1, com relação a classificação de áreas urbanas americanas ou, ainda, na classificação de países segundo o Índice de Desenvolvimento Humano, descrito no Apêndice B.

No caso da classificação de áreas urbanas a dificuldade que se apresentava era a definição da função escore para gerar uma classificação que considerasse o sujeito. Este trabalho mostra como gerar esta classificação considerando a classificação ou escolha dos sujeitos envolvidos. Vimos neste trabalho que para obter a classificação das áreas urbanas para um sujeito, não precisamos saber o quanto ele barganha entre a saúde e a oportunidade de emprego, por exemplo, mas simplesmente observar, uma quantidade razoável de vezes, que cidade ele escolhe, entre várias oportunidades que se apresentam.

Contudo, no caso do IDH, o problema é maior, pois o seu cálculo supõe implicitamente que existem características básicas e suficientes para classificar países, segundo um desenvolvimento humano, sem considerar os sujeitos a que se destina este desenvolvimento.

Outro campo interessante de aplicação deste trabalho é a área de Recursos Humanos de uma empresa. A contratação de um empregado, envolve todos os aspectos aqui explorados. Inicialmente, necessitamos definir os critérios que os candidatos serão julgados, a quantificação dos mesmos e a definição dos pesos de cada um deles. A quantificação destes critérios e a definição dos seus pesos é uma tarefa que constantemente é negligenciada através de uma abordagem subjetiva simplista. Vimos neste trabalho, que a solução é possível, simplesmente oferecendo uma quantidade suficiente de candidatos (reais ou virtuais) ao setor de recrutamento e, observando que escolhas ele faz entre este e aquele candidato. O modelo capta a lógica da escolha e gera a classificação real implícita neste processo, que por outro caminho seria impossível ser gerada.

REFERÊNCIAS BIBLIOGRÁFICAS

1) BADEN-FULLER, C. ; RAVAZZOLO, F. And SCHWEIZER, T. **Making and Measuring Reputations – The Research Ranking of European Business Schools** *Long Range Planning*, vol.33, 621-650, 2000, Elsevier Sciense Ltd.
2) BARBOSA, Sonia Regina da Cal Seixas **Qualidade de Vida e suas Metáforas: Uma Reflexão Sócio-Ambiental** *Tese de Doutorado* 1996 IFCH UNICAMP Campinas-SP.
3) BARNETT, Vic **Discussion at the Meeting on 'Alternatives to economic statistics as indicators of national well-being'** *Journal Royal Statistical Society A* , 161, Part 3, 303-311, 1998.
4) BARROSO, C.L.M. ; MELLO G.N. & FARIA, A. L.G. **Influência de Característica do Aluno na Avaliação do Seu Desempenho** *Fundação Carlos Chagas Caderno de Pesquisa* (26) 61-80, Set 1978
5) BARROSO, C.L.M. ; RIBEIRO NETTO, A ; COELHO M.H.M. **Estudos de Predição do Comportamento Acadêmico I – Faculdade de Medicina Veterinária da USP** *Fundação Carlos Chagas Caderno de Pesquisa* (5) 37-53, Nov 1972
6) BARROSO, C.L.M. ; RIBEIRO NETTO, A ; COELHO M.H.M. **Estudos de Predição do Comportamento Acadêmico I – Faculdade de Medicina da USP** *Fundação Carlos Chagas Caderno de Pesquisa* (5) 55-76, Nov 1972
7) BARROSO, C.L.M. **Pesos Nominais e Pesos Efetivos no Vestibular do CESCEM** *Fundação Carlos Chagas Caderno de Pesquisa* (6) 5-12 , 1972
8) BECKER, R.A. ; Lorraine DENBY; Robert McGILLS and Allan WILKS **Analysis of Data from The Places Rated Almanac** *The American Statistician*, August 1987, vol. 41 nº 3 ASA pg 169-186
9) BICKEL, P.J. and DOKSUM, K.A. **Mathematical Statistics: basic ideas and selected** *Prentice Hall* N.J. 2001.
10) BLOOMFIELD, Peter and STEIGER, William L. **Least Absolute Deviations: Theory , Applications and Algorithms** *Birkhäuser. Boston* 1983.

11) BOX, George ; HUNTER, William and HUNTER, J. Stuart **Statistics for Experimenters** *Wiley & Sons* 1978.

12) BREEN III, T.F. **Estabilidade do Concurso Vestibular do CESCEM** *Fundação Carlos Chagas Caderno de Pesquisa* (12) 49-53, Mar 1975

13) BUCHWEITZ, B. **Testes de Múltipla Escolha e de Resposta Livre em Física Geral** *Fundação Carlos Chagas Caderno de Pesquisa* (16) 3-6, Dez 1976

14) BURTON, R.F. & MILLER, D.J. **Statistical Modelling of Multiple-choice and True/False Tests: ways of considering, and of reducing, the uncertainties attributable to guessing** *Assessment & Evaluation in Higher Education*, vol. 24 nº 4 , 1999 , 399-411

15) CHRISTEN, J. Andrés and NAKAMURA, Miguel **On the Analysis of Accumulation Curves** *Biometrics* 56, 748-754, September 2000.

16) CONWAY, H. McKinley and LISTON, Linda L. **The Good Life Index** *Conway Publications*, Atlanta, USA 1981.

17) COX, D. R. ; FITZPATRICK, D.J. ; FLETCHER, A.E. ; GORE, S.M. ; SPIEGELHALTER, D.J. and JONES, D.R. **Quality-of-life Assessment: Can We keep It Simple?** *Journal Royal Statistical Society A* , 155, Part 3, 353-393, 1992.

18) CUSTANCE, John and HILLIER, Hilary **Statistical issues in developing indicators of sustainable development** *Journal Royal Statistical Society A* , 161, Part 3, 281-290, 1998.

19) DIENER, E.D. and SUH **Measuring Quality of Life: Economic, Social and Subjective Indicators.** *Social Indicators Research* , nº 40, pg 189-216. 1997. Kluwer Academic Publishers.

20) DIENER, ED **A Value Based Index for Measuring National Quality of Life** *Social Indicators Research* , nº 36, pg 107-127. 1995. Kluwer Academic Publishers.

21) EFRON, B and TIBSHIRANI, R.J. **An Introduction to the Bootstrap** *Chapman & Hall/CRC* Washington D.C. 2000.

22) GATTI, B. A . **Vestibular e Ensino Superior nos Anos 70 e 80** *Fundação Carlos Chagas Caderno de Pesquisa* (80) 87-90, Fev 1992

23) GEHRLEIN, William V. and LEPELLEY, Dominique **The probability that all weighted scoring rules elect the same winner** *Economics Letters* 66, 191-197, 2000

24) GILBERT, Nigel **Computer Simulation of Social Processes** *Social Research Update*, March 1993. University of Surrey.

25) GOPEN, George D. and SWAN, Judith A. **The Science of Scientific Writing** *The American Scientist* , vol 78, 550-558, Nov-Dec 1990.

26) HARVEY, Edward B. ; BLAKELY, John H. and TEPPERMAN, Lorne **Toward an Index of Gender Equality** *Social Indicators Research* , nº 22, pg 299-317. 1990. Kluwer Academic Publishers.

27) JOHNSON, R.A. and WICHERN, D.W. **Applied Multivariate Statistical Analysis Third Edition** *Prentice Hall* New Jersey 1992.

28) KACAPYR, Elir **The Well -Being Index** *American Demographics*. Vol. 18 , 2, 32-35, February 1996.

29) KRUSKAL, J.B. and WISH, M. **Multidimensional Scaling** *Series Quantitative Applications in The Social Sciences* . SAGE University Papers, 14th Edition 1989 USA.

30) LAPOINTE, F.J. and LEGENDRE, P. **A Classification of Pure Malt Scotch Whiskies** *Journal Royal Statistical Society D* , vol. 43, n.1 , 237-257, 1994.

31) LEMOS, M.B.; ESTEVES, O. A.; SIMÕES, R. F.; **Uma metodologia para construção de um índice de Qualidade de Vida Urbana** . *Nova Economia*, B.H. vol. 5 nº 2 , Dez 95.

32) LERNER, Sally **Indicators of Human Well Being: Fine -Tuning vs. Taking Action?** *Social Indicators Research* , nº 40, pg 217-220. 1997. Kluwer Academic Publishers.

33) LEVETT, Roger **Sustainability Indicators- Integrating Quality of Life and Environmental Protection** *Journal Royal Statistical Society A* (1998) vol. 161 part 3 pp 291-302.

34) LIU, B. **Social Quality of Life Indicators for Small Metropolitan Areas in America** *International Journal of Social Economics* vol.3 1976

35) **Manual do Candidato Vestibular Nacional UNICAMP 95-99**

36) MCGILL, Robert ; TUCKEY, J.W. ; LARSEN, W. **Variations of Box Plots** *The American Statisticians* vol. 32 nº 1 pp. 12-16 1978.

37) MENDONÇA, D. and RAGHAVACHARI, M. **Comparing the efficacy of ranking methods for multiple round-robin tournaments** *European Journal of Operational Research* 123, 593-605, 2000

38) MICHALOS , Alex C. **Combining Social, Economic and Environmental Indicators to Measure Sustainable Human Well - Being** *Social Indicators Research* , nº 40, pg 221-258. 1997. Kluwer Academic Publishers.

39) MILLAR, J.S. and C. HULL **Measuring Human Wellness** *Social Indicators Research*, nº 40, pg 147-158. 1997. Kluwer Academic Publishers.

40) RIBEIRO NETTO, A. **O Vestibular no Sistema Educacional Brasileiro** *Fundação Carlos Chagas Caderno de Pesquisa* (24) 47-56, Fev 1978

41) RIBEIRO, S.C. ; PESSOA, D. ; KLEIN, R. ; UCHÔA, C.E.F. e FONTANIVE,N.S. **Flutuação de Critérios na Avaliação de Redações** *Educação e Seleção , Fundação Carlos Chagas,* Jul-Dez 1981 nº 4

42) SAVAGEAU, David and LOFTUS, GEOFFREY, Loftus **Places Rated Almanac 5th Edition** *Macmillan* New York , USA 1997.

43) SIMÕES, R.F. ; NAHAS, M.I.; MARTINS, V.L.A . B. e ESTEVES, O . A . **O índice de Qualidade de Vida Urbana de B.H. (IQVU/BH) Como instrumento de Gestão Municipal, produção e Elaboração de Novos indicadores.** *III Conferência Nacional de Geografia e Cartografia* 27 a 31/05/96.

44) **Social Statistics: Follow-Up to the World Summit for Social Development UNCHS**. New York 16-19 April 1996 *Working Group on International Statistical Programmes and Coordination.*

45) **Studies on The Quality of Life in Italy** *Social Indicators Research* , nº 44, volume 1 May 1998. Kluwer Academic Publishers.

46) VIANNA, H.M **Avaliação Educacional nos Cadernos de Pesquisa** *Fundação Carlos Chagas Caderno de Pesquisa* (80) 100-103 , fev 1992

47) VIANNA, H.M **Flutuações de Julgamentos em Provas de Redação** *Fundação Carlos Chagas Caderno de Pesquisa* (19) 5-9, Dez 1976

48) VIANNA, H.M **Processos Alternativos de Seleção para Ingresso no Ensino Superior** *Fundação Carlos Chagas Caderno de Pesquisa* (34) 35-37,Ago 1980

49) VIANNA, H.M **Redação e Medida da Expressão Escrita: Algumas Contribuições da Pesquisa Educacional** *Fundação Carlos Chagas Caderno de Pesquisa* (16) 41-47, Dez 1976

50) VIANNA, H.M. **Testes em Educação** *IBRASA* São Paulo 1973

APÊNDICE A – A CLASSIFICAÇÃO DAS ÁREAS METROPOLITANAS NORTE AMERICANAS SEGUNDO O ÍNDICE DE QUALIDADE DE VIDA URBANA

INVENTÁRIO DE PREFERÊNCIAS

Decida para cada item numerado abaixo, o que seria prioridade para você ao escolher um lugar para morar. Não deixe nenhum item de fora e faça somente uma escolha.

1. E. □ A quantidade de dias quentes (temperatura acima de 32º C) por ano.
   Ou
   A. □ IPTU, taxa de iluminação pública, taxa de consumo de água, eletricidade e esgoto.
2. F. □ O número de assassinatos
   Ou
   A. □ Variedade no oferecimento de escolas.
3. B. □ Qualidade do transporte público
   Ou
   H. □ Quantidade de médicos especialistas.
4. G. □ A quantidade de livros novos adicionados às bibliotecas públicas da cidade
   Ou
   E. □ Condições do clima local, como: umidade, temperatura, quantidade de chuva. . .
5. A. □ O custo da alimentação e do vestuário
   Ou
   B. □ O tempo perdido no deslocamento entre o trabalho e a residência.
6. G. □ A existência de museus de arte e bibliotecas
   Ou
   I. □ A existência de praças de esporte públicas.
7. H. □ A qualidade do ar ao longo do ano
   Ou
   E. □ Risco de temporais e chuvas de granizo.
8. A. □ O preço médio de venda das casas
   Ou
   F. □ Os índices de crimes contra a propriedade
9. C. □ A previsão de crescimento de emprego local
   Ou
   D. □ Uma grande proporção de professores por aluno nas escolas locais
10. G. □ A existência de museus e teatros particulares
    Ou
    F. □ Os índices de arrombamento de carro por ano.
11. I. □ A existência de cursos públicos gratuitos de esportes (futebol, vôlei, natação, etc.)
    Ou
    B. □ Congestionamento nas auto-estradas que cercam a cidade
12. H. □ A existência de Cuidados médicos emergenciais
    Ou
    G. □ A existência de canais de TV e rádios culturais locais
13. G. □ A existência de museus de arte e bibliotecas
    Ou
    A. □ O custo da moradia
14. A. □ O custo da alimentação e do vestuário
    Ou
    C. □ A probabilidade de crescimento do desemprego
15. H. □ A existência de atendimento médico domiciliar público
    Ou
    D. □ O tamanho das escolas públicas
16. G. □ A existência de canais de TV e rádios culturais locais
    Ou

C. □ Oportunidades de emprego no setor de serviço

17. C. □ A taxa de desemprego local

Ou

E. □ O número de dias por ano com tempo bom

18. I. □ Cinemas e bons restaurantes

Ou

D. □ Variedade de escolas públicas e privadas

19. B. □ Transporte público em grande quantidade

Ou

A. □ O preço médio das residências

20. A. □ Os impostos estaduais e municipais sobre a circulação de mercadorias

Ou

H. □ A existência de hospitais escola

21. A. □ O custo com a saúde no local

Ou

I. □ A existência de cursos públicos gratuitos de esportes (futebol, vôlei, natação, etc.)

22. G. □ A existência de canais de TV e rádios culturais locais

Ou

B. □ A existência de aeroporto local e estradas interestaduais

23. H. □ A existência de atendimento médico domiciliar público

Ou

I. □ Cinemas e bons restaurantes

24. F. □ A taxa de crimes violentos

Ou

E. □ A quantidade de dias com tempo bom ao longo do ano

25. D. □ Uma grande proporção de professores por aluno nas escolas locais

Ou

E. □ A quantidade de dias com tempo bom ao longo do ano

26. I. □ Existência de times locais de esporte profissional (Vôlei, Basquete, Futebol,...)

Ou

F. □ A quantidade de arrombamentos e assaltos

27. H. □ Existência de hospitais associados a faculdades de medicina

Ou

F. □ O número de arrombamentos ao longo do ano

28. D. □ Número suficiente de escolas públicas para atender a demanda

Ou

G. □ Companhias de dança e teatro profissional

29. I. □ Proximidade de Parques Aquáticos

Ou

C. □ Novos empregos no setor de manufaturados para o ano que vem.

30. I. □ Proximidade de Parques Nacionais e Florestas

Ou

E. □ Número de dias com tempestade e tormenta ao longo do ano

31. C. □ Empregos de direção em empresas e companhias

Ou

F. □ Número de Arrombamentos ao longo do ano

32. B. □ Serviço aéreo e aeroporto local

Ou

F. □ Número de carros arrombados ao longo do ano

33. B. □ Existência de metro, ônibus e trens

Ou

D. □ Colégios de ensino médio e universidades

34. A. □ Os impostos estaduais e municipais sobre a circulação de mercadorias

Ou

H. □ Existência de hospitais gerais e atendimento médico domiciliar

35. D. □ Os impostos estaduais e municipais sobre a circulação de mercadorias

Ou

A. □ Custo dos serviços públicos e as taxas de propriedade (IPTU, ...)

36. B. □ A existência de aeroporto local e estradas interestaduais

Ou

E. □ Quão frio o inverno pode chegar

37. H. □ Médicos especialistas em todas as áreas

Ou

E. □ A quantidade de dias com bom tempo ao longo do ano.

38. F. □ Número de arrombamentos de casas e carros

Ou

A. □ O custo da moradia

39. G. □ A existência de canais de TV e rádios culturais locais

Ou

I. □ A existência de Zoológicos e parques de entretenimento familiar

40. F. □ O número de assaltos por pessoa no local

Ou

D. □ Uma grande proporção de professores por aluno nas escolas locais

41. C. □ A probabilidade de crescimento de emprego

Ou

B. □ O tempo médio diário desperdiçado em ir ao trabalho e voltar

42. E. □ A variação da sazonalidade da temperatura

Ou

A. □ As taxas sobre a propriedade (IPTU, ...)

43. H. □ A existência de faculdades médicas e hospitais escola

Ou

G. □ Operas e orquestras sinfônicas

44. I. □ A existência de casas de jogo (tipo Bingo)

Ou

B. □ Congestionamento nas auto-estradas que cercam a cidade

45. C. □ Empregos de direção em empresas e companhias

Ou

D. □ Existência de boas escolas particulares

46. E. □ A variação da sazonalidade da temperatura

Ou

C. □ Previsão de crescimento do emprego local

47. F. □ Os índices de arrombamentos de carros, assaltos à mão armada e espancamentos

Ou

E. □ Número de dias com baixíssimas temperaturas

48. A. □ Os custos com climatização da casa

Ou

D. □ Variedade de boas escolas particulares

49. C. □ Expectativa de crescimento nos empregos de alto escalão de companhias e empresas

Ou

F. □ Taxa anual de crimes

50. C. □ O número de novos empregos criados no último ano

Ou

A. □ Os impostos estaduais e municipais sobre a circulação de mercadorias

51. G. □ Rádio com programação exclusiva de música clássica

Ou

B. □ Congestionamento nas auto-estradas que cercam a cidade

52. H. □ Somente cuidados médicos básicos

Ou

I. □ Proximidade de parques nacionais e florestas

53. D. □ Variedade de boas escolas particulares

Ou

E. □ Condições locais de clima, como: umidade, temperatura, quantidade de chuva. . .

54. A. □ O custo da alimentação e do vestuário

Ou

C. □ Empregos na industria

55. E. □ A quantidade de dias quentes por ano.

Ou

B. □ Transporte público em grande quantidade

56. I. □ Boliches, cinemas, restaurantes diversos, . . .

Ou

D. □ Variedade de escolas públicas e privadas

57. G. □ O número de livros em bibliotecas públicas

Ou

C. □ A ameaça de desemprego

58. I. □ Times de Esportes Profissionais (futebol,. . .)

Ou

E. □ Número de dias com baixíssimas temperaturas

59. G. □ Existência de Óperas e Sinfonias

Ou

E. □ Quão frio é o inverno

60. A. □ O preço médio das residências

Ou

G. □ A programação local de artes

61. F. □ A taxa de crimes violentos

Ou

G. □ A variedade da programação de artes

62. F. □ Assaltos e arrombamentos de carros

Ou

H. □ Cuidados médicos especializados

63. C. □ Possibilidade de emprego de agora até o próximo ano

Ou

B. □ Congestionamento nas auto-estradas que cercam a cidade

64. F. □ O índice de crimes contra a propriedade

Ou

I. □ A proximidade de parques nacionais e florestas

65. H. □ Cuidados médicos especializados suficientes

Ou

C. □ Previsão de crescimento de empregos de direção ou alto escalão de empresas e companhias

66. D. □ Boas instituições de ensino superior

Ou

G. □ Número de bibliotecas públicas

67. B. □ Acesso a estradas interestaduais

Ou

D. □ Boas escolas particulares de ensino médio

68. I. □ Cinemas, teatros e bons restaurantes

Ou

A. □ O custo da alimentação e do vestuário

69. H. □ O número de médicos com atendimento familiar

Ou

D. □ A possibilidade de escolher a escola pública de ensino fundamental e médio, mais próxima de sua casa para seus filhos estudarem

70. B. Existência de serviço aéreo na região

Ou

F. Os índices de arrombamento de casas e carros

71. H. □ Qualidade dos médicos e hospitais

Ou

B. □ Qualidade do transporte público

72. H. □ Existência de Especialistas médicos que vêem os pacientes

Ou

C. □ Perspectivas de crescimento de empregos de direção em empresas e companhias.

PREFERÊNCIAS

A = CUSTO DE MORADIA	
B = TRANSPORTE	
C = EMPREGO	
D = EDUCAÇÃO	
E = CLIMA	
F = SEGURANÇA	
G = ARTES	
H = SAÚDE	
I = LAZER	

DADOS BRUTOS

Encontra-se a seguir, o escore combinado para cada um dos nove fatores utilizados no Índice de Qualidade de Vida Urbana para áreas metropolitanas dos Estados Unidos.

Tabela A 1: Escore dos fatores utilizados na classificação das áreas metropolitanas norte americanas segundo o Índice de Qualidade de Vida Urbana

	Metro Area	Moradia	Transporte	Emprego	Educação	Clima	Crime	Artes	Saúde	Lazer
1.	Orange County, CA	2.9	84.5	100.0	95.8	96.8	54.0	95.0	88.2	87.3
2.	Seattle - Bellevue - Everett, WA	9.0	83.4	99.7	88.9	89.0	44.4	96.9	91.5	96.5
3.	Houston, TX	51.4	87.7	100.0	71.2	74.4	25.7	93.2	95.6	94.3
4.	Washington, DC - MD - VA - WV	5.7	95.5	99.7	95.3	55.5	38.5	99.9	97.3	97.9
5.	Phoenix - Mesa, AZ	25.4	85.0	99.9	87.5	81.9	24.3	91.0	85.9	95.3
6.	Minneapolis - ST.Paul , MN - WI	27.5	95.2	100.0	90.0	10.8	64.1	98.0	93.3	96.9
7.	Atlanta, GA	42.6	92.5	100.0	86.1	65.2	11.9	96.0	88.9	92.4
8.	Tampa - St. Petersburg - Clearwater, FL	53.4	87.0	99.4	80.2	78.8	3.1	88.8	86.5	96.4
9.	San Diego, CA	4.9	87.3	100.0	86.9	96.9	20.0	90.6	85.2	96.3
10.	Philadelphia, PA - NJ	15.1	98.4	67.8	94.9	49.4	51.1	99.2	99.6	91.0
11.	San Jose, CA	1.2	85.3	90.0	94.5	93.8	65.1	95.8	65.8	71.8
12.	Long Island, NY	1.4	71.0	65.4	95.3	54.5	85.4	98.8	93.0	97.9
13.	Riverside - San Bernardino, CA	22.4	76.8	99.8	86.9	94.0	8.0	91.3	81.2	96.9
14.	Pittsburgh, PA	34.4	95.1	45.0	85.8	33.2	83.5	92.7	96.6	90.2
15.	Toronto, ON	3.8	98.3	82.9	71.4	25.4	71.6	99.7	98.6	97.7
16.	Portland - Vancouver, OR - WA	11.5	85.1	98.9	87.8	80.0	36.9	85.5	72.3	90.8
17.	Oakland, CA	3.3	80.9	94.3	97.2	96.8	10.5	97.5	75.2	89.3
18.	Denver, CO	22.8	95.0	98.9	92.2	40.3	36.6	90.7	76.3	89.0
19.	Cincinnati, OH-KY-IN	29.7	91.6	85.4	77.8	38.1	58.4	92.3	80.2	86.5
20.	San Francisco, CA	1.1	86.3	55.6	96.6	98.1	13.4	98.6	92.3	96.6
21.	Detroit, MI	27.7	92.7	94.6	91.7	28.3	15.9	96.6	93.6	97.1
22.	Dallas, TX	43.9	62.2	100.0	90.0	62.9	5.0	93.3	84.7	95.3
23.	Chicago, IL	10.0	98.9	91.5	98.2	24.7	5.0	99.8	99.9	99.0
24.	Miami, FL	22.4	84.6	73.3	81.3	87.5	0.0	91.7	89.4	96.6
25.	Cleveland - Lorain - Elyria, OH	20.8	91.4	38.8	88.7	29.9	65.9	98.3	91.9	98.0
26.	Salt Lake City - Ogden, UT	27.2	93.2	98.8	70.9	29.1	61.8	88.4	56.0	95.6
27.	San Antonio, TX	87.7	81.1	94.3	57.3	74.9	20.1	54.1	80.9	68.6
28.	Milwauke - Waukesha, WI	19.6	90.3	66.3	83.2	24.6	56.9	93.0	84.5	96.5
29.	Orlando, FL	43.2	80.7	99.9	72.2	79.2	7.5	79.0	56.9	96.1
30.	Vancouver, BC	3.1	74.9	71.8	73.2	81.5	19.5	94.6	99.0	94.5
31.	Montreal, PQ	16.9	98.1	71.9	75.4	16.2	48.5	98.2	98.8	81.3
32.	Raleigh - Durham -Chapel Hill, NC	18.4	79.3	99.5	85.7	64.8	45.3	81.0	74.7	56.3
33.	Fort Lauderdale, FL	30.4	86.4	90.7	68.5	87.3	6.6	84.3	63.3	85.6
34.	Los Angeles - Long Beach, CA	6.6	94.3	8.5	97.9	97.7	0.4	99.8	99.9	98.0
35.	New Orleans, LA	72.9	77.6	42.9	71.1	73.3	1.6	73.3	89.6	98.6
36.	Indianapolis, IN	48.6	82.0	84.6	75.5	29.6	40.0	91.0	86.4	62.0
37.	Nashville, TN	51.8	74.4	96.2	83.1	51.6	14.4	70.9	78.3	75.5
38.	Sacramento, CA	20.6	70.1	98.9	72.7	88.0	23.6	68.5	62.9	87.1
39.	Kansas City, MO - KS	45.3	86.0	65.0	91.7	26.2	7.0	87.0	91.3	89.7
40.	Rochester, NY	35.0	73.9	60.1	74.7	29.4	74.9	78.1	65.0	96.0
41.	Richmond - Petersburg, VA	30.0	68.5	79.8	66.6	55.4	49.7	81.8	78.8	74.9
42.	Norfolk-Virginia Beach-Newport News, VA-NC	32.8	67.9	71.0	69.9	64.8	44.7	79.3	49.8	96.7
43.	Syracuse, NY	46.6	78.9	29.6	83.8	25.9	85.2	74.1	59.5	90.6
44.	Ventura, CA	3.4	70.0	78.7	70.3	97.5	71.5	55.6	47.1	79.3
45.	Austin - San Marcos, TX	38.0	70.6	98.6	82.2	73.8	37.9	64.1	37.4	70.2
46.	Oklahoma City, OK	80.0	60.9	60.2	83.6	49.5	20.4	69.8	72.8	74.2
47.	Middlesex - Somerset - Hunterdon, NJ	4.4	91.4	88.4	64.2	41.0	86.4	90.3	53.2	50.6
48.	Fort Worth - Arlington, TX	61.4	45.4	99.4	77.1	62.9	13.3	80.2	48.7	78.5
49.	Boston, MA-NH	3.2	93.8	10.3	99.8	46.5	32.1	99.4	92.4	89.1
50.	Omaha, NE - IA	55.5	77.3	47.4	86.6	21.4	51.4	75.6	82.8	66.5
51.	Columbus, OH	29.8	73.8	88.2	82.7	33.9	38.9	91.7	63.8	61.4
52.	Charlotte-Gastonia-Rock Hill, NC-SC	31.5	87.7	94.9	81.0	64.9	5.3	82.4	35.9	80.0
53.	St. Cloud, MN	61.6	94.4	55.4	74.2	8.0	86.5	79.5	57.2	47.0
54.	Memphis, TN - AR - MS	60.7	82.2	82.3	58.0	56.5	9.6	59.7	72.2	76.7
55.	Buffalo - Niagara Falls, NY	38.0	81.3	35.2	82.2	26.4	39.0	94.0	74.5	86.8
56.	St. Louis, MO - IL	20.8	95.3	70.9	90.0	35.0	25.6	96.0	28.7	94.9
57.	Duluth - Superior, MN - WI	87.8	24.8	37.0	69.0	7.7	88.5	68.3	82.0	90.9
58.	Ann Arbor, MI	21.0	78.2	61.3	94.3	28.4	72.0	57.8	68.6	73.8
59.	Grand Rapids - Muskegon - Holland, MI	51.0	54.9	89.2	78.3	22.6	57.0	70.6	33.3	97.8
60.	Louisville, KY - IN	54.9	69.4	68.4	64.4	43.9	59.1	71.1	79.8	43.5
61.	Albany - Schenectady - Troy, NY	22.9	83.9	52.4	81.0	26.7	77.8	72.1	63.2	74.4
62.	New York, NY	4.3	98.3	0.0	99.7	53.1	0.1	100	99.9	97.3

Metro Area	Moradia	Transporte	Emprego	Educação	Clima	Crime	Artes	Saúde	Lazer
63. Santa Rosa, CA	3.4	65.3	73.8	54.7	93.7	68.3	56.1	73.1	61.2
64. Greensboro - Winston - Salem - High Point, NC	31.9	81.4	85.4	77.5	59.8	38.8	65.6	57.2	50.7
65. Fort Wayne, IN	83.6	45.1	61.6	62.8	25.3	75.9	78.8	65.3	47.9
66. Hartford, CT	16.5	83.2	36.4	80.3	35.5	57.9	95.7	62.9	74.5
67. Harrisburg - Lebanon - Carlisle, PA	47.8	64.8	63.1	78.5	42.4	82.5	43.5	63.2	54.9
68. Honolulu, HI	0.3	28.9	42.7	73.2	97.7	69.8	90.0	50.3	87.5
69. Newark, NJ	4.5	96.6	8.4	93.0	48.5	18.5	97.5	91.4	81.8
70. Daytona Beach, FL	55.2	79.2	55.7	76.0	79.9	28.1	43.7	36.3	85.4
71. Tacoma, WA	31.8	67.4	74.0	51.8	80.8	20.9	79.5	49.6	83.1
72. Sarasota - Bradenton, FL	28.3	82.5	77.5	45.9	83.3	18.2	65.0	48.5	88.5
73. Knoxville, TN	71.1	35.8	75.0	62.0	56.1	25.5	56.3	67.4	87.0
74. Santa Barbara - Santa Maria - Lompoc, CA	4.2	58.5	51.1	72.1	96.7	66.8	49.2	54.9	82.4
75. Providence - Fall River - Warwick, RI - MA	19.7	82.1	32.6	88.0	47.0	71.2	70.8	44.3	79.2
76. Monmouth - Ocean, NJ	6.2	83.4	60.5	53.0	53.9	86.1	82.9	25.9	82.8
77. Scranton - Wilkes - Barre - Hazleton, PA	63.3	35.5	24.4	80.3	41.2	93.5	57.2	75.3	61.4
78. Bellingham, WA	29.6	83.5	44.4	46.8	83.1	78.3	57.0	37.7	71.3
79. Dayton - Springfield, OH	41.6	71.0	39.9	84.9	31.1	52.2	83.6	76.8	49.1
80. Baltimore, MD	15.6	97.9	62.8	95.0	56.8	3.9	98.9	7.7	91.4
81. West Palm Beach - Boca Raton, FL	21.6	65.7	89.2	41.2	84.7	3.0	69.9	60.8	92.3
82. Johnson City - Kingsport - Bristol, TN 1- VA	78.7	12.5	50.1	54.8	58.7	89.7	41.2	78.5	62.2
83. Jacksonville, FL	34.2	70.7	80.7	40.4	75.7	1.0	70.9	61.2	89.9
84. Fort Collins - Loveland, CO	38.1	81.2	60.4	50.5	36.6	88.3	40.1	59.1	69.7
85. Tucson, AZ	33.5	62.5	87.5	44.9	79.6	15.1	56.6	59.3	84.6
86. Birmingham, AL	41.3	62.8	64.1	73.4	61.3	12.5	55.1	82.9	68.9
87. Boulder-Longmont, CO	8.3	79.0	66.4	51.3	50.0	77.5	59.1	60.4	70.0
88. Appleton - Oshkosh - Neenah, WI	62.3	52.5	64.7	51.7	14.6	92.8	47.1	57.4	78.7
89. Albuquerque, NM	23.2	76.7	76.0	67.5	57.9	4.5	69.3	72.4	73.7
90. Olympia, WA	32.4	75.8	48.3	38.5	80.0	84.0	86.6	45.5	29.9
91. Wilmington - Newark, DE - MD	22.7	94.1	50.3	69.3	51.3	69.3	59.8	36.7	66.5
92. Greenville - Spartanburg - Anderson, SC	52.8	48.3	78.0	65.9	66.9	19.8	57.1	57.7	71.7
93. Hickory - Morganton - Lenoir, NC	73.8	74.1	62.3	40.8	61.6	75.8	35.0	35.8	58.4
94. Ottawa - Hull, ON - PQ	11.8	92.7	50.8	56.1	16.1	62.6	74.0	94.0	59.0
95. Worcester, MA - CT	12.7	71.6	35.2	92.8	34.0	83.2	68.4	56.0	62.3
96. Bergen-Passaic, NJ	1.9	87.8	29.4	66.7	39.2	78.2	89.8	63.0	58.4
97. Edmonton, AB	28.5	80.3	45.5	51.2	12.9	37.7	77.1	92.0	85.4
98. Spokane, WA	51.8	67.0	44.4	64.6	50.0	54.4	63.0	76.5	38.1
99. Little Rock - North Little Rock, AR	67.3	59.4	65.2	68.0	54.8	1.5	41.8	85.2	66.6
100. Columbia, SC	63.9	40.4	77.9	76.2	65.8	10.1	55.8	52.6	65.7
101. Melbourne - Titusville - Palm Bay, FL	60.7	58.0	56.3	68.0	82.3	27.2	51.8	17.9	86.1
102. Madison, WI	15.9	49.3	67.2	79.0	16.3	78.3	71.7	76.7	53.7
103. Quebec City, PQ	14.5	86.1	36.5	58.4	11.6	74.3	56.6	97.3	69.1
104. Portland, ME	25.0	41.1	49.5	83.0	34.8	78.9	52.5	58.1	79.1
105. Akron, OH	37.4	78.8	45.7	73.6	29.6	62.6	74.3	28.2	71.6
106. Provo - Orem, UT	40.8	61.8	61.3	71.5	33.7	90.7	50.0	12.0	79.4
107. Grand Forks, ND - MN	92.0	61.2	23.0	61.2	5.8	90.0	58.3	74.4	35.0
108. Eugene - Springfield, OR	37.0	47.1	33.7	53.6	83.1	71.3	47.3	48.9	78.1
109. Jackson, MS	84.0	35.6	53.6	69.7	59.4	21.0	45.2	77.6	53.6
110. Bangor, ME	62.3	47.2	29.0	89.8	26.1	91.8	55.4	36.6	60.3
111. New Heaven - Meriden, CT	12.2	96.7	21.9	91.0	47.1	34.3	81.1	54.0	59.9
112. Roanoke, VA	66.2	33.2	30.6	53.0	58.2	77.2	47.5	80.6	51.3
113. Fargo - Moorhead, ND - MN	76.1	57.1	36.3	65.4	6.1	91.6	75.2	66.0	23.9
114. Saskatoon, SK	34.2	67.0	36.3	87.8	8.1	54.4	72.3	93.8	42.1
115. Mobile, AL	78.5	38.2	53.0	55.4	67.2	27.9	44.3	46.1	85.1
116. Charleston - North Charleston, SC	59.9	19.1	59.8	54.8	71.5	19.2	54.1	65.3	91.0
117. Lincoln, NE	68.5	84.6	48.8	86.0	20.6	42.0	58.4	64.3	21.6
118. Tulsa, OK	69.5	51.7	65.3	35.6	47.0	32.9	55.3	55.2	79.0
119. La Crosse, WI - MN	64.8	50.7	38.3	75.7	16.5	91.8	41.1	67.5	44.7
120. Kalamazoo - Battle Creek, MI	62.3	61.1	53.3	77.7	25.4	31.7	56.0	43.5	80.2
121. Erie, PA	77.8	49.4	31.9	52.5	29.2	84.4	39.4	58.6	66.6
122. Halifax, NS	17.9	77.3	29.1	69.9	42.3	28.4	56.1	97.9	70.3
123. Gainesville, FL	61.3	36.5	48.7	59.5	74.6	0.9	52.4	92.2	61.8
124. Salem, OR	36.8	70.1	45.4	79.0	81.5	66.1	41.0	29.1	38.6
125. Calgary, AB	27.2	83.6	42.6	33.6	17.7	49.8	84.4	82.1	65.7
126. Las Vegas, NV - AZ	27.3	82.4	98.2	22.4	74.3	22.6	46.0	16.9	94.7
127. Huntington - Ashland,WV-KY-OH	93.3	27.9	26.4	59.4	48.2	88.6	37.3	55.4	47.8
128. Des Moines, IA	50.0	54.5	56.1	77.5	18.9	62.9	54.0	53.3	55.5
129. Boise City, ID	42.6	61.3	78.1	49.8	49.1	78.9	39.0	49.8	33.9
130. Fayetteville - Springdale - Rogers, AR	82.5	25.7	68.3	42.2	49.6	89.6	13.6	59.8	51.2
131. El Paso, TX	89.4	64.2	86.6	42.1	72.0	69.6	23.4	9.9	24.8
132. Bridgeport, CT	15.9	91.4	33.8	68.1	47.1	40.9	70.8	58.1	54.7
133. London, ON	12.7	88.9	40.2	70.7	20.7	65.0	63.8	96.4	22.2
134. Toledo, OH	48.3	60.1	26.6	62.8	26.8	40.4	80.8	75.2	57.8
135. Springfield, MO	73.0	26.7	68.0	79.1	38.0	75.6	36.0	46.9	35.5
136. Victoria, BC	3.5	68.3	43.9	25.0	84.9	46.3	58.9	95.5	49.8
137. Lexington, KY	55.4	37.7	61.7	85.7	44.2	36.0	50.6	85.3	18.8
138. Biloxi - Gulfport - Pascagoula, MS	88.5	36.7	58.2	1.3	72.8	67.4	33.5	34.4	81.9

Metro Area	Moradia	Transporte	Emprego	Educação	Clima	Crime	Artes	Saúde	Lazer
139. Biloxi - Gulfport - Pascagoula, MS	88.5	36.7	58.2	1.3	72.8	67.4	33.5	34.4	81.9
140. Hamilton, ON	10.5	93.4	32.7	52.8	23.5	66.1	77.8	88.4	29.4
141. Charlottesville, VA	28.2	23.9	39.7	68.2	56.9	84.0	53.5	87.8	30.2
142. Allentown - Bethlehem - Easton, PA	26.8	60.2	34.5	72.9	35.9	90.1	51.6	57.4	42.5
143. Galveston - Texas City, TX	74.5	59.8	51.4	50.7	76.7	19.1	27.6	56.2	55.5
144. Utica - Rome, NY	51.7	68.6	19.6	78.6	23.4	91.4	33.5	32.5	69.7
145. Evansville - Henderson, IN - KY	82.5	38.4	38.3	39.1	41.3	62.1	69.0	69.6	27.9
146. Bloomington, IN	66.7	70.8	43.3	58.5	34.8	78.1	44.5	25.5	45.5
147. Augusta - Aiken, GA-SC	70.2	4.9	75.5	44.2	64.9	42.7	46.6	69.7	48.9
148. Bismarck, ND	84.2	35.5	32.2	35.9	8.7	91.6	69.5	68.9	39.8
149. Vallejo - Fairfield - Napa, CA	11.6	52.9	55.5	56.3	90.4	35.0	55.9	41.0	67.4
150. Lynchburg, VA	79.2	25.7	36.5	65.5	59.7	82.5	25.9	37.0	53.6
151. Sheboygan, WI	70.9	87.8	33.9	6.1	28.4	91.0	54.2	34.4	57.2
152. St. John's, NF	20.0	70.2	27.3	15.9	42.3	87.5	30.7	93.2	75.0
153. Asheville, NC	56.7	17.5	42.1	47.1	60.6	73.9	42.2	73.5	47.4
154. Trenton, NJ	15.5	96.5	36.7	87.0	44.5	47.7	73.7	46.8	11.5
155. Portsmouth - Rochester, NH - ME	10.9	43.6	70.9	84.2	34.6	95.1	51.5	14.3	54.0
156. Winnipeg, MB	16.4	85.6	32.5	60.3	3.4	31.9	79.2	86.0	63.6
157. Iowa City, IA	49.8	42.3	37.1	62.7	21.0	75.9	51.4	97.1	21.4
158. Burlington, VT	25.4	48.5	43.3	87.4	18.5	76.0	18.4	65.7	75.0
159. Wichita, KS	65.5	59.5	53.1	50.1	34.2	32.3	47.8	71.4	44.3
160. Fresno, CA	37.5	63.8	64.1	58.3	84.1	6.0	35.8	42.5	64.4
161. South Bend, IN	82.0	54.2	32.6	77.2	27.9	50.7	54.8	54.0	22.7
162. Reno, NV	14.5	74.7	40.3	57.1	66.8	45.3	28.0	54.2	74.2
163. Gary, IN	44.8	92.7	27.7	65.4	25.6	26.0	73.1	50.5	48.7
164. Charleston, WV	66.0	54.6	26.5	54.7	50.3	78.6	48.0	61.1	14.7
165. Salinas, CA	3.1	36.8	58.6	52.8	96.5	47.4	50.1	29.1	79.8
166. Colorado Springs, CO	39.0	62.0	66.1	71.0	37.2	65.0	54.7	8.4	49.9
167. Columbia, MO	49.8	15.5	40.4	85.0	30.8	65.3	53.6	92.5	20.3
168. Sioux Falls, SD	82.8	50.3	44.6	27.3	9.2	86.2	42.5	89.6	18.2
169. Davenport - Moline - Rock - Island, IA - IL	74.2	59.1	32.5	61.7	21.9	50.0	52.5	42.8	56.0
170. Lancaster, PA	31.3	70.1	31.6	59.9	44.7	92.1	31.0	53.6	35.3
171. Santa Cruz - Watsonville, CA	1.9	67.0	46.1	66.1	96.3	44.4	51.1	34.0	41.7
172. Wheeling, WV - OH	94.6	2.9	27.8	57.5	37.4	95.1	44.9	73.6	14.0
173. Canton - Massilon, OH	47.6	76.0	30.3	46.8	29.6	64.7	75.2	29.6	47.7
174. Peoria - Pekin, IL	67.8	48.1	29.8	74.4	21.7	49.2	57.0	41.6	57.4
175. Brazoria, TX	79.8	55.1	53.7	18.8	74.8	83.5	23.8	3.9	51.9
176. Lubbock, TX	92.6	51.1	40.8	47.6	61.3	45.4	21.8	77.3	6.1
177. Hattiesburg, MS	89.0	5.1	34.2	60.0	63.0	85.0	0.1	67.6	39.4
178. Regina, SK	32.9	75.6	30.3	70.6	4.9	24.9	68.0	86.9	46.2
179. Fort Myers - Cape Coral, FL	26.0	50.9	69.8	7.2	81.5	51.7	42.4	21.6	87.1
180. Dutchess County, NY	22.8	80.6	22.0	68.5	35.5	83.4	44.6	28.3	51.0
181. Springfield, MA	17.2	87.7	18.0	93.7	35.6	24.0	67.5	33.6	58.2
182. Cedar Rapids, IA	52.3	49.8	36.2	75.4	18.1	78.1	51.1	53.8	20.7
183. Bloomington - Normal, IL	52.5	79.0	44.8	79.2	24.3	91.2	38.5	13.4	11.2
184. Greeley, CO	42.1	83.1	44.7	65.7	33.0	69.9	41.6	31.9	20.9
185. Kitchener, ON	11.7	93.9	39.7	47.1	23.1	82.5	64.9	58.5	10.9
186. Santa Fe, NM	9.4	65.2	46.0	71.3	53.1	37.6	41.0	61.0	46.1
187. Macon, GA	83.8	2.2	27.9	55.7	66.1	46.2	54.2	47.4	47.0
188. Jamestown, NY	52.2	50.2	22.1	42.6	29.7	88.1	69.6	11.9	63.5
189. Muncie, IN	82.5	71.1	31.0	63.9	31.7	86.3	18.0	36.6	9.0
190. Greenville, NC	75.7	1.9	59.6	51.0	66.2	28.9	55.8	83.7	5.5
191. Lafayette, LA	88.9	14.8	51.5	25.9	68.7	60.4	28.1	37.1	52.3
192. Johnstown, PA	85.6	32.0	31.9	39.9	36.9	95.7	25.2	52.6	27.6
193. Huntsville, AL	39.5	14.0	63.0	80.2	58.2	45.7	27.7	44.4	54.6
194. Tallahassee, FL	39.7	38.7	53.8	68.5	68.7	0.1	27.2	60.6	69.5
195. Stamford - Norwalk, CT	1.7	82.7	33.8	2.4	43.6	80.0	85.9	58.1	38.6
196. Fort Pierce - Port St. Lucie, FL	73.3	17.6	53.1	40.4	84.7	35.5	32.9	9.0	80.2
197. Lawrence,KS	71.3	58.0	31.8	92.3	33.3	58.7	29.2	26.2	25.6
198. Eau Claire, WI	83.1	15.8	37.4	32.9	10.4	91.8	41.5	80.8	31.9
199. Barnstable - Yarmouth, MA	6.5	63.1	35.9	31.3	59.5	44.1	79.7	13.1	92.4
200. Kenosha, WI	54.5	81.4	26.9	65.7	27.1	78.8	36.1	3.5	49.0
201. St. Catharines - Niagara, ON	14.4	73.9	24.4	26.2	30.4	79.0	70.0	67.5	35.3
202. Athens, GA	60.5	89.0	31.0	68.0	64.9	36.3	38.6	28.9	3.7
203. Champaign - Urbana, IL	48.2	80.6	55.4	64.3	25.4	50.9	50.7	44.7	0.6
204. Redding, CA	47.5	24.3	36.1	47.3	81.1	67.0	12.5	53.7	50.4
205. Windsor, ON	16.8	95.6	26.5	31.4	28.3	76.0	49.3	61.8	33.8
206. Manchester, NH	14.4	28.0	33.1	93.1	30.4	85.8	42.1	46.2	46.5
207. Youngstown - Warren, OH	75.1	16.2	28.2	47.6	29.6	80.1	55.8	26.6	60.0
208. Amarillo, TX	93.1	42.1	34.5	33.2	51.7	35.7	36.6	60.2	31.8
209. Lansing - East Lansing,MI	49.8	48.4	49.5	81.5	23.9	55.4	44.8	30.9	34.4
210. Lakeland - Winter Haven, FL	55.7	72.5	36.4	73.3	79.2	5.9	1.6	11.1	79.4
211. Tuscaloosa, AL	78.4	49.9	45.3	83.4	61.7	10.2	18.9	36.5	29.5
212. Pensacola, FL	53.6	43.0	47.7	33.5	73.3	13.4	23.2	51.3	73.9
213. Waterloo - Cedar Falls, IA	94.6	41.4	33.4	37.8	15.3	73.9	40.2	62.5	12.8
214. Reading, PA	39.1	69.4	25.4	60.6	35.6	82.2	31.0	34.6	33.8

Metro Area	Moradia	Transporte	Emprego	Educação	Clima	Crime	Artes	Saúde	Lazer
215. Reading, PA	39.1	69.4	25.4	60.6	35.6	82.2	31.0	34.6	33.8
216. Parkesburg - Marietta, WV - OH	89.9	9.0	23.4	42.0	45.0	93.4	50.2	22.4	36.4
217. Yolo, CA	12.8	74.6	56.4	49.4	87.6	40.6	24.7	52.6	7.8
218. Kokomo, IN	85.8	50.9	27.9	35.1	29.1	79.1	49.0	45.1	4.2
219. San Luis Obispo - Atascadero - Paso Robles, CA	5.8	16.8	46.9	30.6	96.9	74.5	32.2	45.1	57.3
220. Shreveport - Bossier City, LA	89.2	33.5	23.3	19.1	59.6	10.0	39.9	72.9	58.3
221. Killeen - Temple, TX	96.3	17.5	41.9	50.0	66.5	64.3	1.4	43.7	23.0
222. Florence, AL	90.6	0.3	43.3	39.4	56.8	88.2	14.2	30.6	40.5
223. Nashua, NH	12.8	42.5	33.1	63.9	32.5	95.6	47.4	46.2	29.7
224. Bremerton, WA	28.5	49.6	30.4	27.3	82.5	84.0	37.3	18.6	44.6
225. Rochester, MN	42.9	35.9	36.7	32.8	9.4	91.4	49.2	99.8	3.8
226. Green Bay, WI	39.4	53.6	56.4	61.4	15.8	86.4	39.5	19.5	28.6
227. Decatur, AL	82.4	2.4	33.9	44.2	56.8	84.7	27.1	21.4	46.8
228. Punta Gorda, FL	59.0	27.2	57.7	0.5	80.3	89.3	13.9	10.7	61.0
229. Columbus, GA - AL	85.4	20.0	34.1	31.0	66.9	57.3	27.4	40.5	34.9
230. Savannah, GA	43.1	37.6	36.5	40.1	74.9	24.4	42.4	26.6	71.0
231. Corpus Christi, TX	74.3	25.6	40.6	14.7	85.6	20.7	23.9	46.3	62.4
232. Brownsville - Harlingen - San Benito, TX	91.0	21.1	49.8	25.9	88.8	34.7	6.5	2.9	72.5
233. Grand Junction, CO	60.7	25.4	37.4	22.0	34.4	74.6	30.6	69.7	38.2
234. Terre Haute, IN	80.3	3.1	32.7	57.7	30.9	92.4	31.0	46.2	16.4
235. Montgomery, AL	61.7	18.6	41.4	64.7	65.3	34.8	21.6	47.1	34.3
236. Sherbrooke, PQ	14.3	25.7	23.5	80.7	15.9	77.9	22.8	97.2	31.4
237. Springfield, IL	65.4	61.5	26.5	58.6	26.3	27.0	38.7	64.8	20.1
238. Clarcksville - Hopkinsville, TN-KY	95.5	60.0	38.3	49.6	47.1	30.2	54.7	3.3	10.1
239. Saginaw - Bay City - Midland, MI	70.6	20.0	30.1	52.1	25.6	36.2	51.5	45.3	57.0
240. Anchorage, AK	19.2	47.4	57.1	44.1	32.7	30.7	32.1	73.1	51.2
241. St. Joseph, MO	86.3	71.5	20.8	42.2	26.1	65.0	50.2	19.3	6.0
242. Baton Rouge, LA	71.6	14.1	64.3	26.2	68.2	0.0	43.2	38.9	60.7
243. Chattanooga, TN-GA	82.0	16.6	42.5	48.7	57.3	28.2	15.0	39.3	57.6
244. Lafayette, IN	74.8	27.3	46.8	57.2	26.4	83.1	43.6	18.6	7.9
245. Beaumont - Port Arthur, TX	93.8	18.3	29.8	20.2	70.7	16.2	28.7	44.4	62.5
246. Dubuque, IA	86.5	33.0	33.2	51.8	16.1	86.6	43.7	21.6	12.1
247. Benton Harbor, MI	76.4	44.8	30.7	34.7	26.8	11.6	71.8	20.5	65.8
248. Fort Smith, AR - OK	97.5	2.9	54.3	1.7	51.5	74.2	3.9	50.3	46.4
249. Houma, LA	94.7	0.1	28.7	3.2	74.7	68.2	16.4	13.9	82.9
250. Danbury, CT	6.7	65.0	33.8	15.3	36.6	91.9	54.0	58.1	20.4
251. Jackson, TN	80.3	5.3	37.0	58.9	52.0	2.4	58.2	84.1	2.8
252. Chico-Paradise, CA	30.5	17.9	39.9	41.8	85.1	66.9	16.9	42.3	38.1
253. Williamsport, PA	78.2	19.7	22.3	30.6	32.9	92.8	28.4	69.0	3.9
254. Lawrence, MA - NH	17.4	57.4	22.4	65.2	36.1	48.4	59.5	19.0	52.0
255. Alexandria, LA	97.0	9.2	32.4	13.7	68.9	11.4	40.4	64.1	39.9
256. Flint, MI	60.5	84.9	22.2	71.6	26.7	6.9	41.9	31.6	29.4
257. Fitchburg - Leominster, MA	14.8	57.9	35.2	59.2	33.9	34.6	49.4	56.0	33.0
258. Medford - Ashland, OR	42.7	22.2	43.5	2.5	73.7	69.1	40.1	31.2	47.7
259. Waco, TX	94.1	31.0	33.2	53.3	64.1	14.5	19.4	40.1	21.5
260. Thunder Bay, ON	13.3	59.3	23.4	23.2	7.5	39.3	41.0	74.1	88.3
261. Janesville - Beloit, WI	85.4	27.5	24.9	47.0	18.9	78.9	52.8	13.4	20.0
262. Oshawa, ON	14.7	95.6	34.3	15.4	32.4	83.7	32.0	46.5	14.2
263. Saint John, NB	25.5	32.8	28.6	0.0	27.0	88.1	26.7	65.8	73.9
264. Flagstaff, AZ - UT	24.9	20.6	37.3	23.1	44.3	63.3	28.3	45.9	80.3
265. Lowell, MA - NH	15.8	49.2	22.7	25.5	36.1	44.1	51.2	75.7	47.0
266. Wausau, WI	83.0	30.6	42.5	10.4	12.9	94.4	25.6	44.4	23.4
267. Hamilton - Middletown, OH	38.4	61.9	38.5	57.2	36.0	63.7	47.6	6.9	15.1
268. Elkhart - Goshen, IN	88.0	29.8	53.6	21.3	22.6	82.8	40.2	21.2	4.5
269. Atlantic City - Cape May , NJ	14.1	87.8	40.1	29.9	51.6	10.7	39.2	4.9	84.5
270. Billings, MT	83.5	48.0	30.0	18.9	27.4	77.0	23.2	40.8	13.4
271. Fort Walton Beach, FL	49.5	23.2	43.5	17.7	69.4	82.2	0.0	12.8	62.3
272. Chicoutimi-Jonquiere, PQ	41.7	19.1	15.3	33.4	6.0	86.0	13.8	73.0	71.8
273. Tyler, TX	93.0	3.4	40.1	45.7	68.9	25.9	0.0	57.6	22.7
274. Owensboro, KY	82.3	7.5	29.3	41.9	48.3	88.7	19.2	26.4	13.0
275. McAllen - Edinburg - Mission, TX	96.8	5.4	67.3	8.7	87.9	35.6	18.2	12.4	23.8
276. Naples, FL	10.2	31.9	62.6	6.8	82.0	35.9	32.1	14.7	79.7
277. Bryan College Station, TX	97.1	4.3	53.9	35.6	74.0	50.5	9.6	20.0	9.8
278. New London - Norwich, CT - RI	13.2	27.5	31.5	31.9	45.9	86.8	58.4	9.3	47.2
279. Abilene , TX	93.5	24.3	22.4	34.7	64.0	49.5	30.7	29.1	3.1
280. Wilmington, NC	43.0	12.7	49.8	28.2	69.6	18.4	23.4	22.0	82.7
281. Rocky Mount, NC	68.3	48.2	41.8	44.1	65.7	36.2	27.5	14.7	1.2
282. Sudbury, ON	16.5	54.7	28.4	20.3	8.2	65.1	20.2	69.0	65.0
283. Cheyenne, WY	88.7	29.9	22.3	6.7	30.9	85.4	29.8	50.5	3.2
284. State College, PA	47.3	18.8	38.4	63.2	30.9	93.6	29.3	8.0	17.4
285. Glens Falls, NY	44.0	37.1	26.9	8.9	18.5	80.9	34.7	35.1	59.5
286. Altoona, PA	81.8	20.3	25.9	7.2	34.1	94.7	21.3	60.0	0.5
287. Pittsfield, MA	33.7	37.4	18.6	1.7	28.7	82.6	58.5	37.6	46.2
288. Brockton, MA	15.5	41.8	32.5	74.8	40.0	37.9	37.9	8.3	54.6
289. Longview - Marshall, TX	91.2	4.6	39.9	45.0	62.6	37.1	20.4	21.6	20.6
290. Visalia - Tulare - Porterville, CA	51.0	27.1	38.9	22.5	85.5	38.3	22.3	18.8	38.6

Metro Area	Moradia	Transporte	Emprego	Educação	Clima	Crime	Artes	Saúde	Lazer
291. Rapid City, SD	85.4	20.8	30.7	16.9	24.9	68.2	15.1	38.9	41.4
292. Hagerstown, MD	27.2	74.8	31.3	13.3	38.9	90.3	36.6	3.4	26.2
293. Waterbury, CT	15.4	70.7	21.9	32.5	36.6	59.4	35.0	58.1	12.1
294. Myrtle Beach, SC	50.0	19.9	63.2	10.6	64.1	9.1	23.1	21.4	79.9
295. Las Cruces, NM	60.3	43.5	45.8	42.5	65.7	49.4	0.0	6.0	26.3
296. Newburgh, NY - PA	16.3	38.5	36.9	34.6	39.2	81.3	33.4	4.1	54.8
297. Joplin, MO	91.0	5.3	44.4	40.5	41.8	82.6	10.2	17.9	4.4
298. Binghamton, NY	37.8	35.0	20.6	21.3	26.9	91.1	43.9	18.6	40.1
299. Great Falls, MT	87.1	45.8	22.4	0.5	27.9	61.0	20.9	42.4	26.4
300. Pueblo, CO	61.0	59.6	28.6	35.4	36.7	5.3	35.0	56.5	15.8
301. Sherman - Denison, TX	96.8	0.4	29.3	14.0	54.4	58.8	25.2	28.8	25.2
302. Steubenville - Weirton, OH - WV	95.2	1.5	17.7	41.2	36.7	89.2	34.1	6.4	10.6
303. Pine Bluff, AR	96.8	39.5	25.1	13.9	56.8	10.8	17.3	61.3	11.0
304. Monroe, LA	94.7	20.1	29.0	9.1	55.3	17.4	29.9	56.4	20.4
305. San Angelo, TX	96.3	16.6	28.9	6.0	66.0	47.7	31.4	24.8	13.6
306. Bakersfield, CA	44.6	29.5	49.1	25.2	85.7	19.5	24.5	10.6	42.1
307. Richland - Kennewick - Pasco, WA	53.0	24.5	38.3	2.3	68.5	70.8	25.3	13.3	34.5
308. Topeka, KS	89.1	85.7	30.7	12.7	28.9	6.2	44.8	18.4	13.1
309. Cumberland, MD - WV	64.1	4.8	21.9	40.6	42.6	86.9	26.4	36.7	4.9
310. Victoria, TX	97.0	1.4	33.5	19.8	76.3	17.6	15.0	64.3	3.0
311. Trois - Rivieres, PQ	29.6	3.7	25.3	41.8	12.0	79.6	49.4	76.4	9.5
312. York, PA	43.4	22.0	44.1	25.1	44.3	91.4	18.1	21.6	16.7
313. Danville, VA	95.5	10.9	21.6	19.3	58.1	91.4	20.0	1.9	3.3
314. Yakima, WA	61.4	19.9	30.8	6.2	57.7	29.9	31.3	36.5	47.3
315. Laredo, TX	92.0	16.2	77.3	18.3	84.6	25.3	1.6	0.1	2.9
316. Dothan, AL	90.1	3.5	40.6	28.4	59.9	43.7	11.1	37.8	2.6
317. Sharon, PA	82.6	6.5	24.5	40.1	33.7	92.6	1.7	14.4	21.4
318. Florence, SC	91.8	11.4	31.9	27.4	66.8	11.9	13.0	59.7	2.8
319. Merced, CA	43.1	24.1	37.3	8.7	84.4	61.8	17.3	16.4	23.6
320. Modesto, CA	36.1	22.8	48.8	24.8	85.8	19.0	27.4	35.5	15.3
321. Wichita Falls, TX	97.9	21.2	34.5	3.7	53.7	21.1	17.1	47.4	18.6
322. Goldsboro, NC	78.3	43.3	27.8	32.0	65.9	41.2	13.8	10.1	2.3
323. Anniston, AL	96.3	5.4	30.9	24.0	62.0	25.4	36.0	16.1	18.6
324. Fayetteville, NC	85.5	27.8	38.4	30.9	65.4	8.0	33.7	15.1	9.9
325. Lewiston - Auburn, ME	50.5	5.6	27.4	34.4	30.4	90.0	16.3	40.4	17.1
326. Odessa -Midland, TX	79.9	23.4	44.0	25.0	68.6	31.2	21.6	10.5	6.8
327. Casper, WY	99.2	11.0	23.0	4.9	22.8	54.8	41.5	37.9	15.6
328. Racine, WI	55.6	50.6	27.7	0.0	24.5	58.8	39.3	6.3	47.7
329. Jersey City, NJ	11.0	74.0	25.6	37.9	47.2	7.2	62.4	24.8	20.1
330. Jacksonville, NC	75.8	3.3	31.4	1.5	69.4	81.3	9.3	1.4	35.8
331. Gadsden, AL	94.3	17.8	32.5	21.3	58.1	16.5	33.5	26.2	8.3
332. Sioux City, IA - NE	51.3	39.3	29.6	31.6	16.2	28.9	37.3	58.6	14.4
333. Lake Charles, LA	93.2	11.1	29.6	3.9	69.0	17.6	17.7	48.0	16.6
334. Texarkana, TX - Texarkana, AR	99.3	4.7	28.8	19.0	63.3	35.4	0.7	42.3	12.1
335. Kankakee, IL	66.2	70.3	31.8	40.5	23.3	24.3	9.1	28.9	10.9
336. Ocala, FL	76.4	2.3	63.2	1.5	77.1	10.6	14.0	2.6	56.2
337. Panama City, FL	62.6	4.8	35.5	12.2	74.2	29.9	15.7	4.9	64.0
338. Decatur, IL	69.9	17.9	13.3	29.4	25.6	44.9	42.0	53.6	5.7
339. Rockford, IL	49.6	21.5	30.8	10.6	19.4	37.7	42.1	54.1	30.4
340. New Bedford, MA	27.4	68.8	26.0	18.0	51.8	21.9	36.0	1.6	44.3
341. Stockton - Lodi, CA	21.7	6.0	42.7	36.8	88.6	8.7	38.5	20.9	27.4
342. Yuba City, CA	49.2	32.6	33.2	20.0	86.2	27.8	15.2	6.5	18.0
343. Enid, OK	99.1	6.5	22.8	0.3	44.3	34.5	13.7	54.6	0.9
344. Yuma, AZ	53.3	17.2	40.5	5.0	89.3	35.2	8.0	0.6	27.2
345. Albany, GA	79.2	11.2	25.2	27.7	68.9	9.3	35.5	10.0	7.5
346. Sumter, SC	82.2	5.2	32.2	50.0	70.4	8.9	11.7	0.1	10.3
347. Lawton, OK	95.3	9.9	31.8	5.4	52.4	40.5	1.0	16.9	17.3
348. Lima, OH	50.6	6.0	27.9	28.2	27.6	24.2	46.0	43.8	13.3
349. Jackson, MI	53.3	21.5	24.1	32.6	25.6	29.3	30.6	1.5	48.8
350. Vineland - Millville - Bridgeton, NJ	20.4	78.5	22.5	3.7	52.1	21.3	36.4	13.3	15.6
351. Dover, DE	36.5	2.7	36.8	36.2	56.8	58.3	4.1	0.0	24.8
352. Elmira, NY	54.2	26.7	25.8	5.8	29.7	8.3	44.6	26.0	5.9
353. Mansfield, OH	51.9	4.6	22.4	6.2	29.6	35.4	50.5	11.3	8.5
354. Visalia - Tulare - Porterville, CA	51.0	27.1	38.9	22.5	85.5	38.3	22.3	18.8	38.6
355. Rapid City, SD	85.4	20.8	30.7	16.9	24.9	68.2	15.1	38.9	41.4

APÊNDICE B – A CLASSIFICAÇÃO DAS NAÇÕES SEGUNDO O ÍNDICE DE DESENVOLVIMENTO HUMANO(IDH)

INTRODUÇÃO

Segundo o Relatório do Desenvolvimento Humano de 2002, o objetivo principal do Relatório é avaliar o estado do desenvolvimento humano em todo o mundo e fornecer, em cada ano, uma análise crítica de um tema específico. Para maiores informações sobre o último relatório, acesse o seguinte endereço na Rede Mundial de Computadores http://www.undp.org.br/FTP/HDR2002/ .

O Índice de Desenvolvimento Humano (IDH) é utilizado para classificação das nações (em 2002, foram analisadas 173 nações) e, é uma medida resumo do desenvolvimento humano através da realização média de um país em três dimensões básicas:

- Uma vida longa e saudável, medida pela esperança de vida ao nascer;
- Conhecimento, medido pela taxa de alfabetização de adultos (com ponderação de dois terços) e pela taxa de escolarização bruta combinada do primário, secundário e superior (com ponderação de um terço) e
- Um nível de vida digno, medido pelo PIB per capita (dólares PPC)

CALCULANDO O IDH

Antes de calcular o próprio IDH, é necessário criar um índice para cada uma destas três dimensões. Para o cálculo destas indicadores de dimensão – índices de esperança de vida, educação e PIB – são escolhidos valores mínimos e máximos (baliza) para cada indicador primário.

Desta forma, o desempenho de cada dimensão é expresso como um valor entre 0 e 1, utilizando a seguinte fórmula geral:

$$Índice\ de\ Dimensão = \frac{valor\ atual - valor\ mínimo}{valor\ atual - valor\ máximo}$$

O IDH é então calculado como uma média simples dos índices de cada dimensão. Os valores mínimos e máximos utilizados em cada dimensão encontram-se na tabela abaixo.

Tabela B 1: Balizas para o cálculo do IDH

Indicador	Valor Mínimo	Valor Máximo
Esperança de vida ao nascer (em anos)	85	25
Taxa de Alfabetização de adultos (em %)	100	0
Taxa de escolarização bruta combinada (em %)	100	0
PIB per capita (dólares PPC)	40.000	100

Vejamos, como exemplo o cálculo do IDH para o Brasil.

1. CÁLCULO DO ÍNDICE DA ESPERANÇA DE VIDA

$$Índice\ da\ esperança\ de\ vida = \frac{67.7 - 25}{85 - 25} = 0.71$$

2. CÁLCULO DO ÍNDICE DA EDUCAÇÃO

$$Índice\ de\ alfabetização\ de\ adultos\ (\%15\ anos\ ou\ mais) = \frac{85.2 - 0}{100 - 0} = 0.852$$

$$Índice\ de\ escolarização\ bruta = \frac{80 - 0}{100 - 0} = 0.80$$

$$Índice\ da\ Educação = \frac{2}{3} Índice\ de\ alfabetização\ de\ adultos + \frac{1}{3} Índice\ de\ escolarização\ bruta =$$

$$= \frac{2}{3} 0.852 + \frac{1}{3} 0.80 = 0.835 \approx 0.83$$

3. CÁLCULO DO ÍNDICE DO PIB

O rendimento é ajustado porque para alcançar um nível elevado de desenvolvimento humano não é necessário um rendimento ilimitado.

$$\textit{Índice do PIB} = \frac{\log(7.625) - \log(100)}{\log(40.000) - \log(100)} = \frac{3.88 - 2}{4.60 - 2} = \frac{1.88}{2.60} = 0.72$$

4. CÁLCULO DO IDH

$$\textit{IDH} = \frac{1}{3}\,\textit{índice da esperança de vida} + \frac{1}{3}\,\textit{índice da educação} + \frac{1}{3}\,\textit{índice do PIB} =$$

$$= \frac{1}{3}0.71 + \frac{1}{3}0.835 + \frac{1}{3}0.723 = 0.757$$