



XIX COBREAP | Foz do Iguaçu

INOVAÇÕES CIENTÍFICAS E TECNOLÓGICAS

AGOSTO 2017

**Quando utilizar RNA,
Envoltória de Dados e
Regressão espacial ?**

Engenheiro Mecânico e Civil:

Eng. Antonio Pelli Neto

Orientador Mestrado:

Prof. Dr. Antônio de Pádua Braga

**Aplicação para o município
de Belo Horizonte / MG**

Agosto 2017

Resumo

- A construção de modelos de formação de preços imobiliários através das técnicas convencionais, como é atualmente realizada, enfrenta problemas que diminuem a precisão das estimativas de valores, especialmente pelo desconhecimento da forma funcional que descreve a relação entre as variáveis e pela dificuldade da estimação dos parâmetros relativos à distribuição espacial dos imóveis. A aplicação das redes neurais artificiais em avaliação de imóveis apresenta boas perspectivas e os resultados obtidos até o momento indicam que as RNA's são ferramentas computacionais melhores para áreas onde os dados não podem ser representados unicamente por modelos lineares.

Organização da Apresentação

- Capítulo 1 – Introdução
- Capítulo 2 – Mercado Imobiliário e o Mercado de habitação (contextualização)
- Capítulo 3 – Metodologias aplicadas às avaliações imobiliárias
- Capítulo 4 – Desenvolvimento (estudo de caso para a cidade de Belo Horizonte)
- Capítulo 5 – Conclusões e perspectivas futuras

Introdução – Importância das Avaliações

- Mercado Imobiliário
 - Geração de empregos na construção civil
 - Arrecadação de impostos e tributos
- Valor de mercado:
 - Elaboração de plantas de valores genéricos, cobrança de impostos sobre ganhos de capital, desapropriações, decisões do Poder Judiciário, garantia de operações comerciais, análise de viabilidade econômica e operações de compra e venda de imóveis.

Formação dos Preços

- Preços dos imóveis
 - Modelos hedônicos (estruturais, locacionais, econômicas e temporais)

$$P_{(x)} = p(x_1, x_2, \dots, x_n)$$

- Valor de mercado
 - Estimado com base em uma amostra de preços do mercado imobiliário
 - Método usual:
 - Regressão linear múltipla (MQO)
 - Modelos Clássicos de Regressão Linear (MCR).

Motivação

- Dificuldades de aplicação do MCR
 - Autocorrelação espacial
 - Desconhecimento da forma funcional
- NBR 5676
 - Fatores de homogeneização
 - Regressão Linear
- NBR 14.653 – 2 (tratamento dos dados, Avaliar)
 - Fatores de homogeneização
 - Regressão Linear
 - Redes Neurais Artificiais
 - Regressão Espacial
 - DEA – Análise da Envoltória de Dados sob Dupla Ótica
 - Outros métodos de indução de valor

Objetivos

- Primeiro objetivo:
 - Demonstrar a possibilidade da utilização das Redes Neurais Artificiais na Engenharia de Avaliações:
 - Avaliações em massa (RNA x Regressão Espacial), normalmente séries temporais.
 - Avaliações individuais (EDO – não é objeto deste estudo), dados de corte.
- Segundo objetivo:
 - Elaborar uma análise microeconômica do mercado imobiliário, com base no tratamento escolhido.

Mercado Imobiliário

- Base de estudo – Valor de Mercado
 - Mercado habitacional
 - Estudos de demanda
- Aplicação de recursos financeiros
 - 18 bilhões de reais em 2006
 - Erros ou imprecisões indesejáveis podem afetar expressiva parcela da população

Histórico

- 1974 – I Congresso Brasileiro de Avaliações e Perícias em Engenharia – COBREAP
- 1989 – NBR 5676 (Grande impulso desapropriações do Governo Federal de Fernando Collor de Melo)
- 1998 – Avaliar (Simpósio eminentemente técnico)
- 2004 – Parte 2 NBR 14.653 (tratamento técnico dos dados de mercado)

Mercado Imobiliário

- Enfoques básicos
 - Renda
 - Custo
 - Comparação
- NBR 14.653 – 1
 - Opção pelo Método Comparativo de Dados de Mercado.
 - Tratamentos técnicos (RNA's, Regressão Espacial, Regressão Linear, DEA, outros).

Método Comparativo

- Variáveis dependentes, explicadas, de saída ou output:
 - Valor total
 - Valor unitário
- Variáveis independentes, explicativas, entradas ou inputs:
 - Características intrínsecas e extrínsecas
 - Estruturais, econômicas e locais (macrolocalizativas ou microlocalizativas).

Variáveis independentes

- Quantitativas
- Dummy ou binárias
- Proxy
- Qualitativas
 - Ausências de estudos específicos, como aplicação de algoritmos de separação em classes
 - KNN, K-Médias, Perceptron, MLP

Metodologias aplicáveis

- Regressão Linear Múltipla
 - Mínimos quadrados ordinários, validação inferência estatística
- Regressão Espacial
 - Metodologia proposta por Anselin
- Redes Neurais Artificiais
 - MLP – aproximação de função

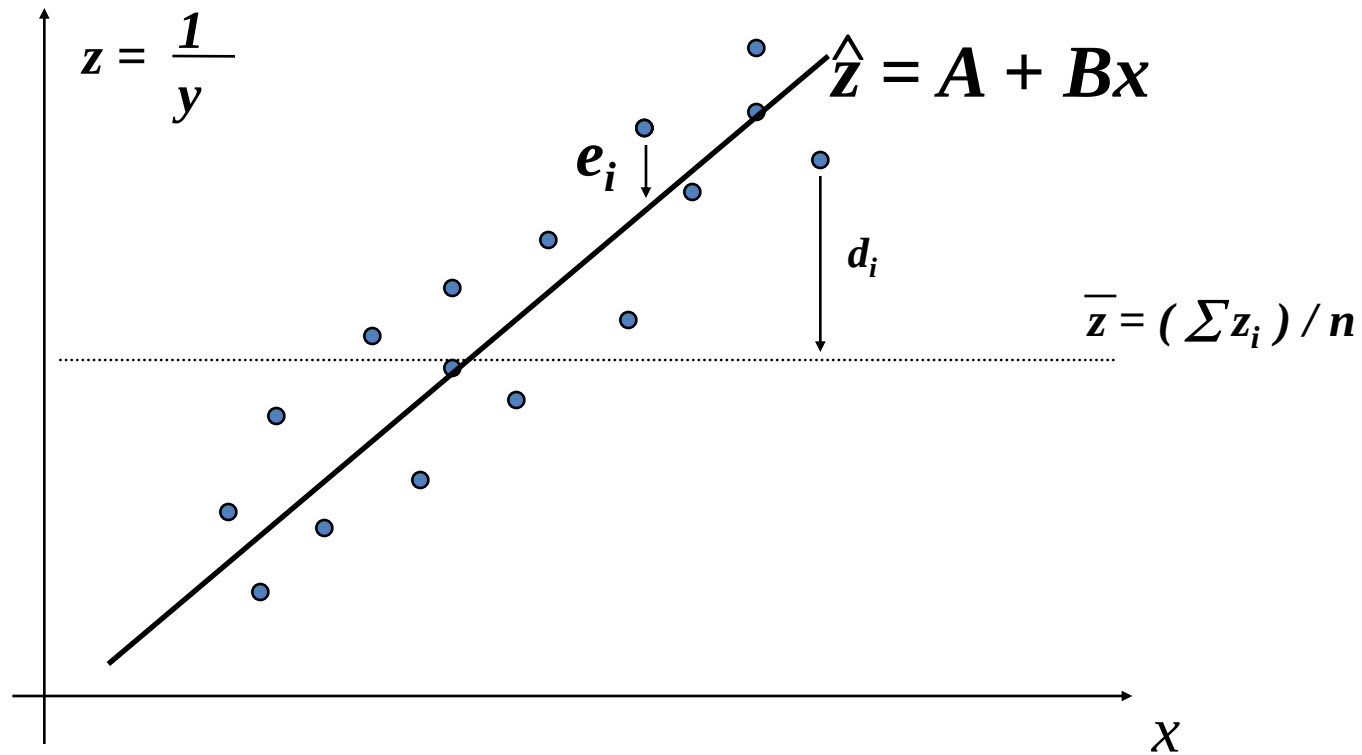
Regressão Linear Múltipla

- Metodologia tradicional (preços hedônicos)

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i, i = 1, \dots, m$$

- onde Y_i é a variável dependente, explicada ou variável saída, $Y_1, \dots, Y_m$ são as variáveis independentes, explicativas ou de entrada, normalmente associadas às características físicas, de localização, e aos aspectos econômicos, $X_{i1}, \dots, X_{ik}$ são denominados de parâmetros da população e ε_i são os erros aleatórios do modelo.
- Forma funcional do modelo $\beta_0, \dots, \beta_k$ é desconhecida $\varepsilon_1, \dots, \varepsilon_m$
 - solução: transformação de variáveis

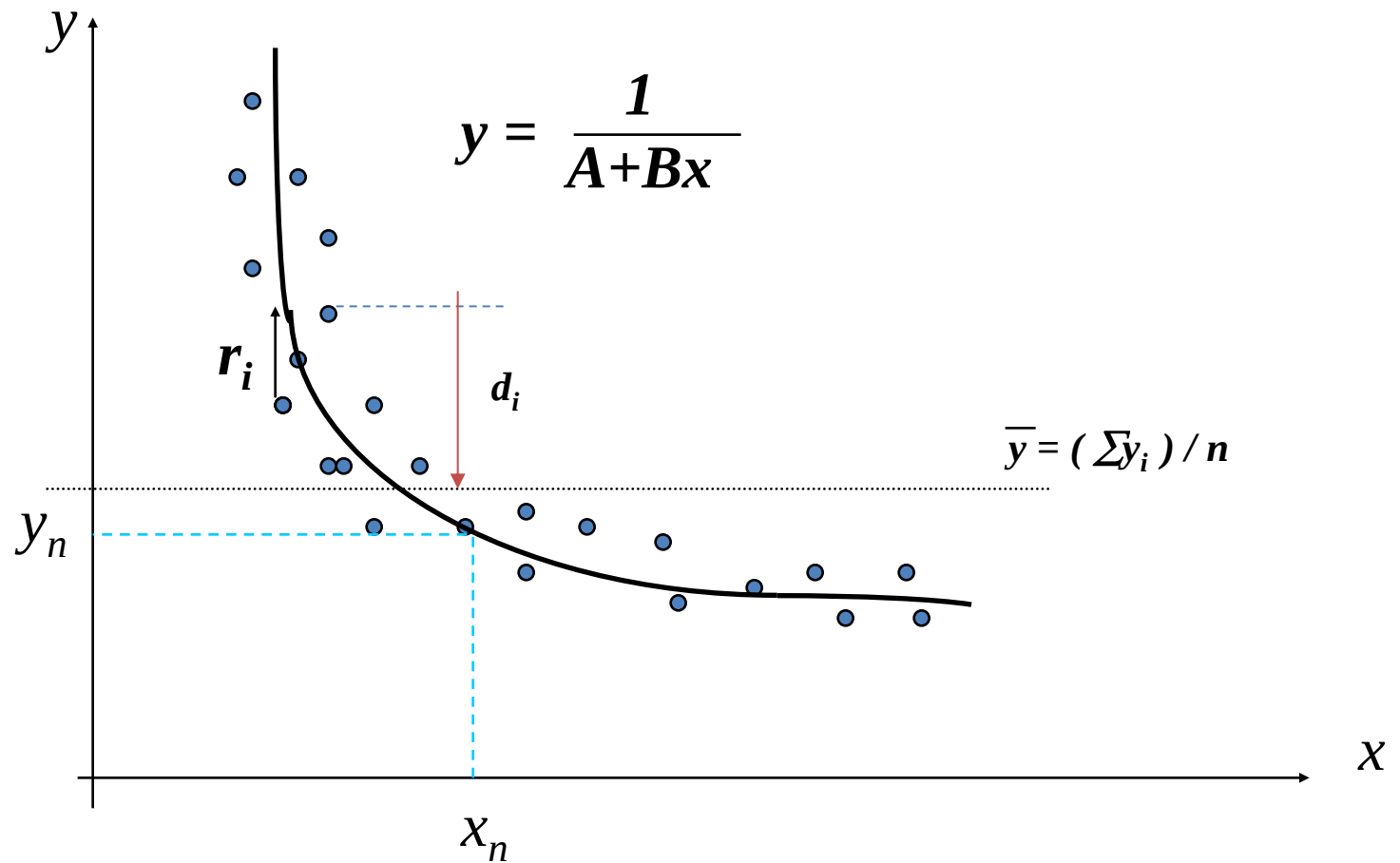
Transformações variáveis



$$z = \frac{1}{y} \rightarrow y = \frac{1}{A+Bx}$$

$$C_d = \frac{\sum d_i^2 - \sum e_i^2}{\sum d_i^2}$$

Transformações variáveis



Regressão Linear Múltipla

- Ruptura das hipóteses básicas
 - Normalidade
 - Homocedasticidade
 - Não autoregressão
 - Colinearidade
 - Pressuposto de linearidade (transf. Variáveis)
- Testes estatísticos inconsistentes

Regressão Espacial (Anselin 1998)

- Características Locacionais:
 - Macrolocalizativas – explica a variação da média do bairro em relação à média da cidade
 - Microlocalizativas – explica a variação do valor do bem em relação à média do bairro

Regressão Espacial

- Formas de Considerar as Características Macrolocalizativas
 - Distâncias a Pólos de Influência
 - Divisão da Região em Cidades, Bairros, etc.
 - Coordenadas Geográficas dos Imóveis
- Variáveis introduzidas nos modelos tradicionais

Regressão Espacial

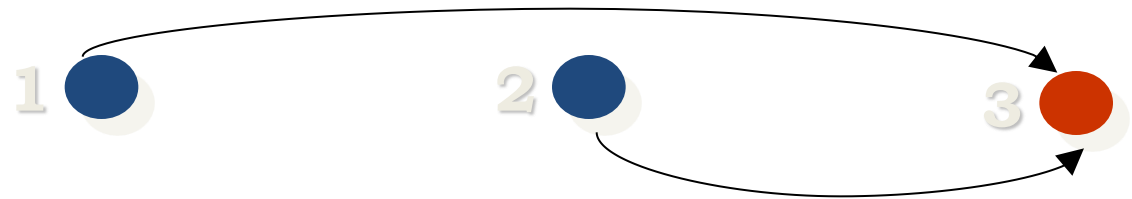
- Formas de Considerar as Características Microlocalizativas
 - Posição relativa de cada imóvel em relação aos seus vizinhos, de forma que cada um sofre uma influência maior daqueles que estão mais próximos a ele.
- Nova abordagem: Regressão Espacial
 - Introduzida no Brasil pelo Eng. Rubens Alves Dantas

Regressão Espacial

- “...dados associados à posição que ocupam no espaço (cidades, regiões, bairros ou a própria coordenada geográfica), são caracterizados pela dependência ou heterogeneidade espacial (Anselin, 1988).”

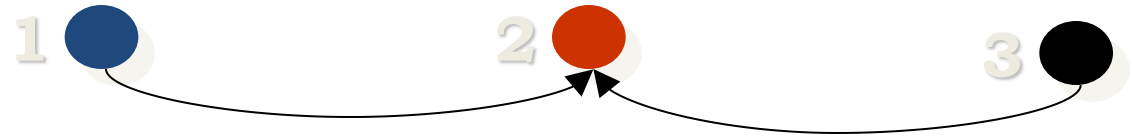
Interação Espacial

**Polo
valorizante**



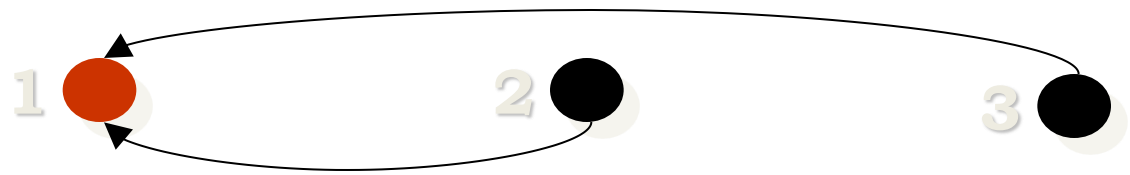
Interação Espacial

**Polo
valorizante**



Interação Espacial

**Polo
valorizante**

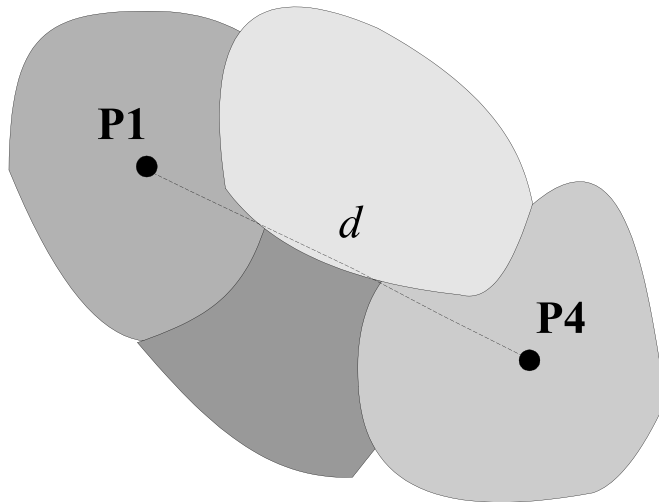


Regressão Espacial

- 1º. passo: Diagnóstico da Dependência Espacial
 - Análise Gráfica do Variograma
 - Testes estatísticos específicos
 - Construção matriz de espacialidade
- 2º. passo: Modelagem
 - Modelos de Dependência Espacial (Regressão Espacial e extensão para as Redes Neurais Artificiais)

Regressão Espacial

- Matriz de espacialidade



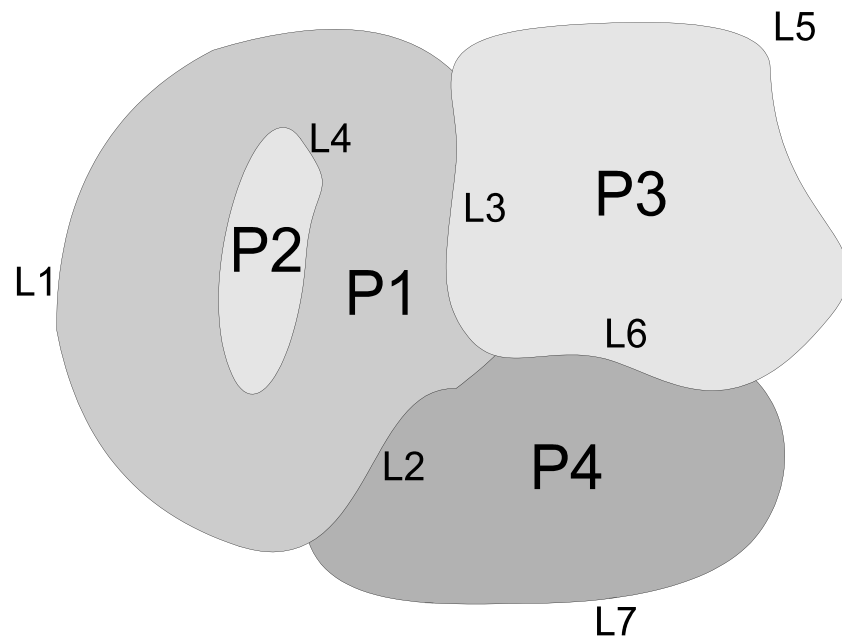
$$w_{14} = 0 \quad \text{para } d > \text{limiar}$$

$$w_{14} = 1 \quad \text{para } d \leq \text{limiar}$$

- Inverso da distância linear.

$$w_{14} = \frac{1}{d}$$

Matriz de Espacialidade



$$W = \begin{bmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \\ w_{31} & w_{32} & w_{33} & w_{34} \\ w_{41} & w_{42} & w_{43} & w_{44} \end{bmatrix}$$

w_{ij} : “distância” do objeto i ao objeto j .

Regressão Espacial

- Modelo de erro espacial

$$Y = X\beta + (I - \lambda W)^{-1}u$$

- Modelo de defasagem espacial

$$Y = X\beta + \rho WY + \varepsilon$$

Redes Neurais Artificiais

- Trabalhos desenvolvidos por profissionais área
 - **1999**
 - **ESTUDO DA INFLUÊNCIA DE LOTES NA ACESSIBILIDADE NO VALOR URBANOS ATRAVÉS DE REDES NEURAI** - *BRONDINO , Nair Cristina Margarido – Tese de Doutorado na Universidade de São Carlos - SP*
 - **2000**
 - **PLANTA DE VALORES UTILIZANDO INTELIGÊNCIA ARTIFICIAL -** *GONZÁLES Marco Aurélio Stumpf - COBRAC 2000, Congresso Brasileiro de Cadastro Técnico Multifinalitário, UFSC Florianópolis .*
 - **APLICAÇÃO DE REDES NEURAI NA AVALIAÇÃO DE BENS** - Uma Comparação com a Análise de Regressão – *GUEDES, Jackson Carvalho AVALIAR - II Simpósio Brasileiro de Engenharia de Avaliações – GUEDES, Jackson Carvalho – Setembro de 2000 – São Paulo - Pgs. Anais 3 a 10;*
 - **O EMPREGO DE INTELIGÊNCIA ARTIFICIAL NA AVALIAÇÃO DE BENS -** *GUEDES, Jackson Carvalho.*

Redes Neurais Artificiais

- 2001
- **ANÁLISE DA UTILIZAÇÃO DE INTELIGÊNCIA ARTIFICIAL NA DETERMINAÇÃO DE PLANTAS DE VALORES** - GONZÁLEZ, Marco Aurélio Stumpf (1); FORMOSO, Carlos Torres (2)
- **DUAS FERRAMENTAS PODEROSAS À DISPOSIÇÃO DO ENGENHEIRO DE AVALIAÇÕES - MODELOS LINEARES GENERALIZADOS E REDES NEURAI** - GUEDES, Jackson Carvalho - IBAPE – XI COBREAP – Congresso Brasileiro de engenharia de Avaliações e Perícias - Vitória-ES
- 2002
- **MÍNIMOS QUADRADOS EFICAZES, REDES NEURAI** EFICIENTES - MELO, Brício de– AVALIAR - III Simpósio Brasileiro de Engenharia de Avaliações – Curitiba-PR
- **APLICAÇÃO DE TÉCNICAS DE DESCOBRIMENTO DE CONHECIMENTO EM BASES DE DADOS E DE INTELIGÊNCIA ARTIFICIAL EM AVALIAÇÃO DE IMÓVEIS.** GONZÁLEZ, Marco Aurélio Stumpf - Tese (Doutorado em Engenharia Civil) – Programa de Pós-Graduação em Engenharia Civil, UFRGS, Porto Alegre / RS.

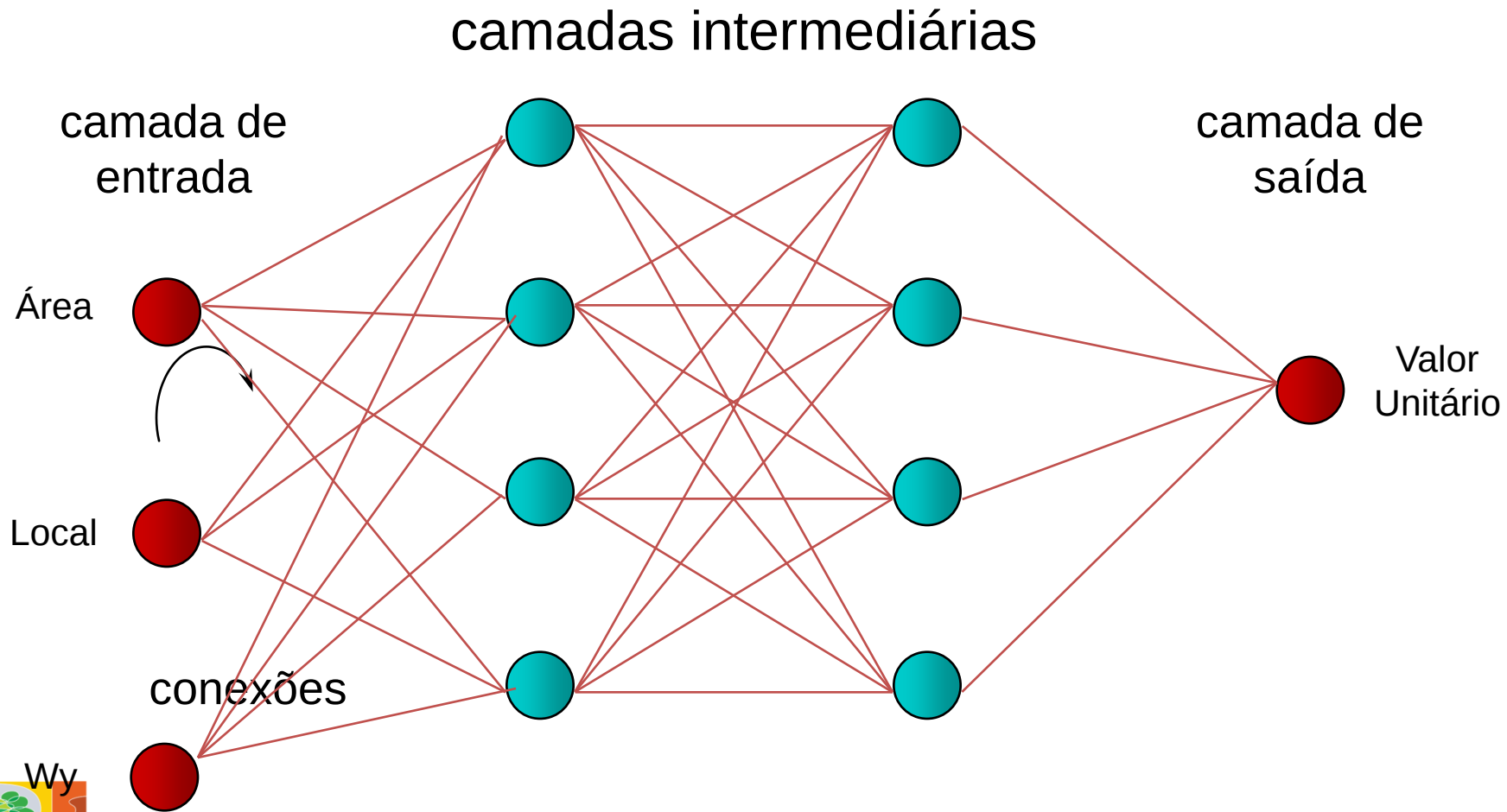
Redes Neurais Artificiais

- 2003
- **AVALIAÇÃO DE IMÓVEIS URBANOS COM UTILIZAÇÃO DE REDES NEURAIAS ARTIFICIAIS –**
PELLI, Antonio Neto ; ZARATE, Luiz - IBAPE-MG – XII COBREAP – Congresso Brasileiro de Engenharia de Avaliações e Perícias – Belo Horizonte – MG
- 2003
- **AVALIAÇÃO DE NAVIOS COM USO DE REGRESSÃO LINEAR E COMPARAÇÃO DOS RESULTADOS COM REDE NEURAL – MORATO, Paulo César Vidal; GUEDES Jackson Carvalho -**
IBAPE-MG – XII COBREAP – Congresso Brasileiro de Engenharia de Avaliações e Perícias – Belo Horizonte – MG
- **ESTUDO COMPARATIVO ENTRE REDES NEURAIAS ARTIFICIAIS E ANÁLISE DE REGRESSÃO MÚLTIPLA NA AVALIAÇÃO DE BENS, PARA PEQUENAS AMOSTRAGENS – SCHIAVO, Eliane Hasselmann Camardella ; AZEVEDO , Márcio Pacheco - IBAPE-MG – XII COBREAP – Congresso Brasileiro de Engenharia de Avaliações e Perícias – Belo Horizonte –MG**

Redes Neurais Artificiais

- 2004
- **AVALIAÇÃO DE IMÓVEIS URBANOS COM UTILIZAÇÃO DE SISTEMAS NEBULOSOS (REDES NEURO-FUZZY) E REDES NEURAS ARTIFICIAIS** *PELLI, Antonio Neto - XXI Congreso Panamericano de Valuación - SCdA UPAV, Cartagena,- Colômbia.*
- **UTILIZAÇÃO DE REDES NEURAS ARTIFICIAIS NA PREVISÃO DE VARIÁVEIS DE FLUXOS DE CAIXA NA AVALIAÇÃO ECONÔMICA DE NEGÓCIOS EM SITUAÇÃO DE RISCO** – Linha de pesquisa - FREIMAN, José Paulo – Universidade Federal de Itajubá – MG.

Redes Neurais Artificiais



Wy

Redes Neurais Artificiais

- Com aplicação de técnicas
 - Prunning
 - Bagging

Estudo de caso – Belo Horizonte

TABELA 4.1 - Área total, população e densidade demográfica
Brasil, MG, RMBH, BH – 2000

Especificação	População	Área (km²)	Densidade
BH	2.238.526	330,90	6.764,96
RMBH	4.357.942	9.459,10	460,71
MG	17.891.494	586.552,40	30,50
BR	169.799.170	8.514.215,30	19,94

Fonte: IBGE. Censo Demográfico 2000

Estudo de caso – Belo Horizonte

TABELA 4.2 - Domicílios em Belo Horizonte por espécie e unidade de planejamento - 2000

Nome UP	Total de domicílios	Domicílios particulares			Domicílios coletivos
		Total de domicílios particulares	Domicílios particulares permanentes	Domicílios particulares improvisados	
Noroeste	97.051	96.107	95.916	191	944
Norte	51.551	51.017	50.780	237	534
Oeste	77.902	77.112	76.949	163	790
Pampulha	40.440	39.768	39.668	100	672
Venda Nova	65.358	65.015	64.896	119	343
Barreiro	70.286	69.953	69.747	206	333
Centro-Sul	84.354	83.001	82.833	168	1.353
Leste	73.279	72.452	72.193	259	827
Nordeste	76.294	75.952	75.465	487	342
Belo Horizonte	636.515	630.377	628.447	1.930	6.138

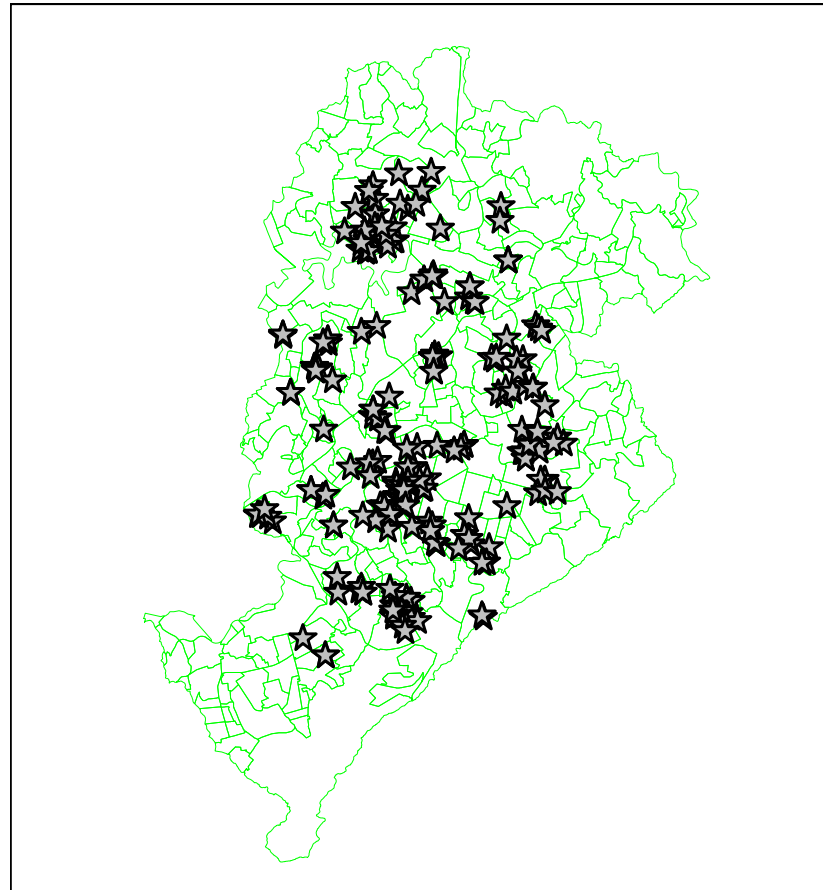
Administração pública



Coleta de dados

- Início em fev./2004 e término set./2005
- Período de coleta – 20 meses
- Todos dados vistoriados
- Banco de dados da Caixa Econômica Federal
- Coordenadas geográficas fornecidas pela Prefeitura Municipal de Belo Horizonte

Coleta de dados



Construção das variáveis

TABELA 4.3 – Descrição da amostra utilizada para estimação dos modelos de preços hedônicos

Variável	Mínimo	Média	Máximo	Amplitude
Setor Urbano	250	484	900	650
Nível / Elevador	5	8	26	21
Data Evento	1	5	20	19
Total de Vagas	0	121	400	400
Área Coberta	40	83	300	260
Área Descoberta	0	5,6	100	100
Dormitórios	2	2	4	2
Equipamentos	1	8	32	31
Unidades p/Acesso	2	11	80	78
Padrão Acabamentos	120	258	600	480
Estado de Conservação	240	473	600	360
Preços Unitários	400,00	960,96	3.333,33	2.933,33

Heterogeneidade

- Os preços de venda variam de 400,00 R\$/m² a 3.333,33 R\$/m² (entre R\$25.000,00 a R\$ 1.000.000,00) – o mais caro tem preço 40 vezes superior ao mais barato. Retirados dados com valores superiores a 2.000 R\$/m².
- As áreas privativas interna dos imóveis também possuem grande variabilidade, sendo a maior área 7,5 vezes superior ao de menor área.
- Os imóveis mais bem localizados estão 3,6 vezes melhor pontuados em relação aos imóveis localizados nos setores urbanos mais populares.
- A área descoberta dos apartamentos varia de 0,00 m² a 100 m², indicando uma variação de 100% entre os imóveis que não possuem área descoberta e aqueles que estão situados nos pavimentos térreos, com áreas privativas descobertas.
- Existem também imóveis recém construídos e até aqueles com idade de 40 anos, com conservação variando de ruim a boa e padrão de mínimo a luxo.

Regressão Linear Múltipla

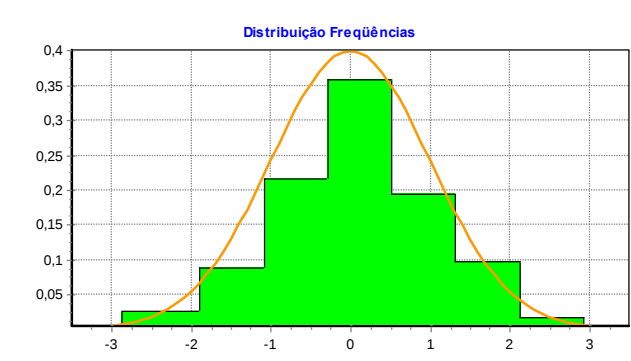
$$P = f(L, E, T, \beta) + \varepsilon$$

TABELA 4.4 – Resultados da estimação dos modelos de preços hedônicos com a regressão linear

Variáveis	Transformação	B (regressor)	t - Student	α (%) (bi caudal)
Setor Urbano	x	+0,0008226328371	8,24	< 0,01
Nível/Elevador	Ln(x)	+0,1613379329	3,50	0,05
Data do Evento	X	+0,005733353163	2,35	1,95
Total Vagas	X	+0,001348528869	6,70	< 0,01
Área Coberta	Ln(X)	-0,5026981916	-8,71	< 0,01
Área Descoberta	X	+0,002915663288	4,42	< 0,01
Dormitórios	X	+0,1168737474	4,70	< 0,01
Equipamentos	X	+0,009908174092	5,31	< 0,01
Unidades p/ Acesso	Ln(X)	-0,05386150912	-3,97	< 0,01
Padrão Acabamento	X	+0,001020533186	6,45	< 0,01
Estado Conservação	X	+0,0009130749008	5,99	< 0,01
Preços Unitários	Ln(Y)	+7,039344981		

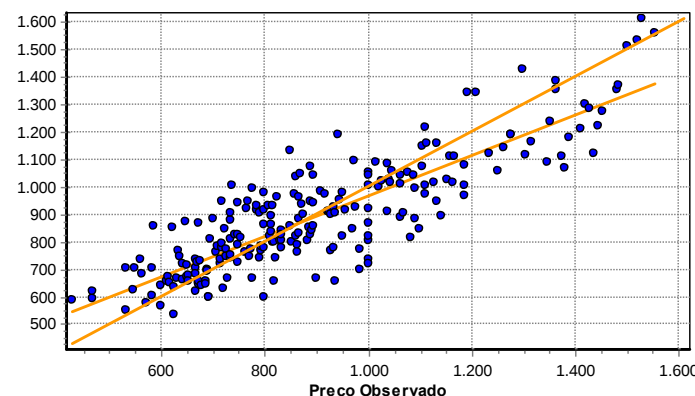
Regressão Linear Múltipla

- Modelo de preços hedônicos
 - Valorização período (20 meses): 11,50%
 - Localização: 50% valor imóvel
- Testes estatísticos
 - F – 53,71 ($\alpha < 0,01\%$)
 - t – $\alpha < 2\%$
 - Normalidade resíduos



Regressão Linear Múltipla

- Coerência hipóteses formuladas
- Modelo Empírico
- Valor Unitário = $e^{(+7,039344981 + 0,0008226328371 * \text{Setor Urbano} + 0,1613379329 * \ln(\text{Nível/Elevador}) + 0,005733353163 * \text{Data do Evento} + 0,001348528869 * \text{Total Vagas} - 0,5026981916 * \ln(\text{Área Coberta}) + 0,002915663288 * \text{Área Desc.} + 0,1168737474 * \text{Dormitórios} + 0,009908174092 * \text{Equipamentos} - 0,05386150912 * \ln(\text{Unidades por Acesso}) + 0,001020533186 * \text{Padrão} + 0,0009130749008 * \text{Conservação})}$
- Preços x
Valores Observados



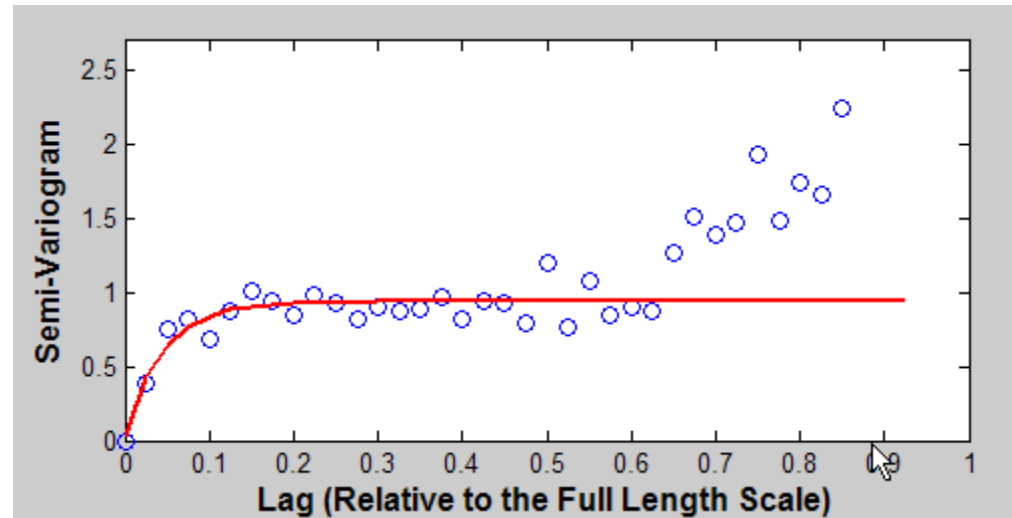
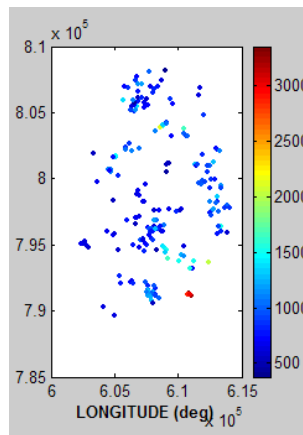
Regressão Espacial

$$P = f(L, E, T, \beta, WP) + \varepsilon$$

- onde f é indicativo da forma funcional, β são os parâmetros do modelo, W a matriz de pesos espacial e ε são os resíduos ou erros aleatórios. Serão utilizados para a modelagem os mesmos dados e transformações realizadas no modelo MCR da seção anterior.

Regressão Espacial

- Construção variograma



- Distância estimada de 2,8 a 3 Km

Regressão Espacial

- Matriz de pesos W (226 x 226)

Dado	1	2	3	4	5	6	7
* 1	0	591,300000	660,110000	2.652,430000	2.505,230000	2.329,680000	2.326,940000
2	591,300000	0	171,250000	2.857,710000	2.695,500000	2.507,260000	2.451,720000
3	660,110000	171,250000	0	2.733,420000	2.568,970000	2.379,310000	2.314,120000
4	2.652,430000	2.857,710000	2.733,420000	0	171,420000	361,890000	517,350000
5	2.505,230000	2.695,500000	2.568,970000	171,420000	0	191,600000	356,520000
6	2.329,680000	2.507,260000	2.379,310000	361,890000	191,600000	0	223,860000
7	2.326,940000	2.451,720000	2.314,120000	517,350000	356,520000	223,860000	0
8	3.706,930000	4.039,690000	3.941,650000	1.373,060000	1.536,420000	1.711,050000	1.890,180000
9	4.202,500000	4.577,130000	4.490,080000	2.000,470000	2.160,870000	2.330,400000	2.516,350000
10	4.202,500000	4.577,130000	4.490,080000	2.000,470000	2.160,870000	2.330,400000	2.516,350000
11	4.341,760000	4.699,300000	4.606,790000	2.050,900000	2.216,370000	2.393,130000	2.568,200000
12	3.541,440000	3.930,780000	3.850,120000	1.521,640000	1.662,000000	1.808,130000	2.015,130000
* 13	3.375,760000	3.713,690000	3.618,120000	1.119,060000	1.272,270000	1.435,190000	1.628,790000
14	3.375,760000	3.713,690000	3.618,120000	1.119,060000	1.272,270000	1.435,190000	1.628,790000
15	3.375,760000	3.713,690000	3.618,120000	1.119,060000	1.272,270000	1.435,190000	1.628,790000
16	3.336,730000	3.681,190000	3.587,600000	1.121,410000	1.270,590000	1.428,900000	1.626,730000
17	3.681,140000	4.038,140000	3.947,090000	1.461,610000	1.617,730000	1.782,790000	1.974,160000

Regressão Espacial

- Testes de autocorrelação espacial

**Tabela 4.5 Diagnóstico da Dependência Espacial
de preços por habitação**

Teste MI/GL	Valor	Probabilidade
LM Robusto (erro)	5,73	0,01665
LM Robusto (defasagem)	16,7544	0,00004

- Modelo adotado – Defasagem espacial

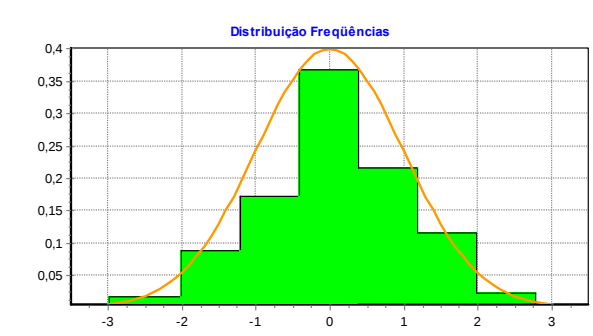
Regressão Espacial

TABELA 4.6 – Resultados da estimação dos modelos de preços hedônicos com a regressão linear

Variáveis	Transformação	B (regressor)	t - Student	α (%) (bi caudal)
Setor Urbano	x	0,000706624	7,2	< 0,01
Nível/Elevador	Ln(x)	0,178771353	4,10	<0,01
Data do Evento	X	0,005513866	2,40	1,62
Total Vagas	X	0,001296873	6,83	< 0,01
Área Coberta	Ln(X)	-0,506194192	-9,33	< 0,01
Área Descoberta	X	0,002978795	4,80	< 0,01
Dormitórios	X	0,120864298	5,17	< 0,01
Equipamentos	X	0,009361433	5,33	< 0,01
Unidades p/ Acesso	Ln(X)	-0,05063036	-3,97	< 0,01
Padrão Acabamento	X	0,000956406	6,37	< 0,01
Estado Conservação	X	0,00082779	5,74	< 0,01
WY	X	0,262942164	3,696	0,02
Preços Unitários	Ln(Y)	+5,34366799		

Regressão Espacial

- Modelo de preços hedônicos
 - Valorização período (20 meses): 11,16%
 - Localização: 40% valor imóvel
- Testes estatísticos
 - F – 56,25 ($\alpha < 0,01\%$)
 - t – $\alpha < 2\%$ ($W_y = 0,02\%$)
 - Normalidade resíduos



Regressão Espacial

- Coerência hipóteses formuladas
- Modelo Empírico

Valor Unitário = $e^{(+5,343664799 + 0,0007066240486 * \text{Setor Urbano}$

$+ 0,1787713532 * \ln(\text{Nível/Elevador}) + 0,00551386571 * \text{Data do Evento} + 0,00129687329 * \text{Total Vagas} - 0,5061941919$

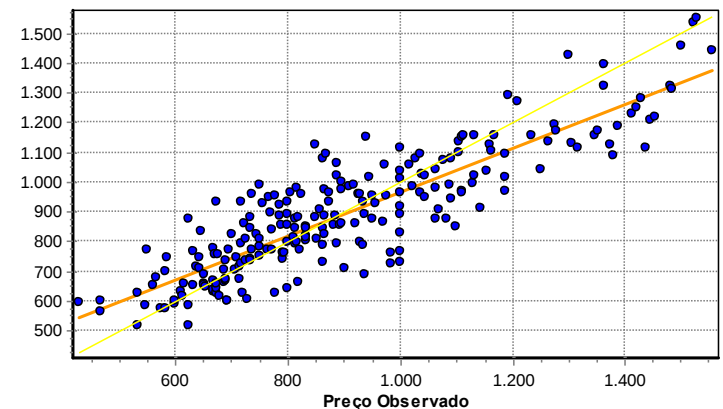
$* \ln(\text{Área Coberta}) + 0,002978794592 * \text{Área Descoberta}$

$+ 0,1208642976 * \text{Dormitórios} + 0,009361433193 * \text{Equipamentos}$

$- 0,05063036011 * \ln(\text{Unidades por Acesso}) + 0,0009564062885 * \text{Padrão} + 0,0008277902485 * \text{Conservação}$

$+ 0,2629421637 * w * y$

- Preços x
- Valores Observados



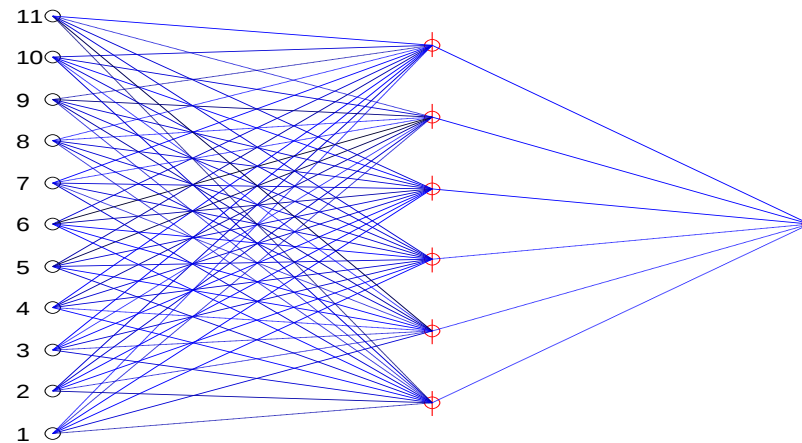
Redes Neurais Artificiais

$$P = f(L, E, T, w) + \varepsilon$$

- onde f é indicativo da forma funcional, w os pesos a serem ajustados durante a fase de treinamento da rede e ε são os resíduos ou erros aleatórios. Serão utilizados para a modelagem os mesmos dados, porém sem transformações, do modelo MCR

Redes Neurais Artificiais

Arquitetura da Rede 5



- Algoritmo de treinamento *Levenberg Marquadt*
- Estrutura rede: empírica (6 neurônios camada escondida)
- Função de transferência (tangente hiperbólica camada escondida, linear na camada de saída)
- Épocas (de 15 a 50), com erro quadrático mínimo 0,00001.

Redes Neurais Artificiais

TABELA 4.7 – Resultados da estimação com as RNA's

Variáveis	Valor Médio	Variação (%)
Setor Urbano	484,3142	3,98
Nível/Elevador	8,0575	3,56
Data do Evento	5,3407	0,99
Total Vagas	115,0442	4,04
Área Coberta	80,5880	-8,18
Área Descoberta	5,4513	5,62
Dormitórios	2,7566	2,04
Equipamentos	8,4735	2,65
Unidades p/ Acesso	11,4513	-1,80
Padrão Acabamento	250,7080	4,70
Estado Conservação	473,8938	2,88
Preço Unitário Estimado	876,88	

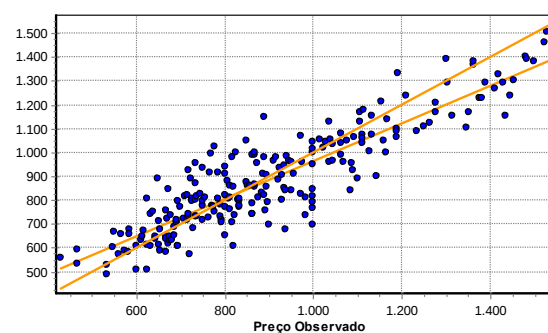
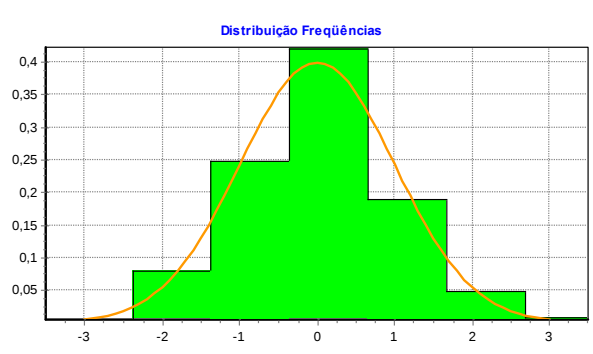
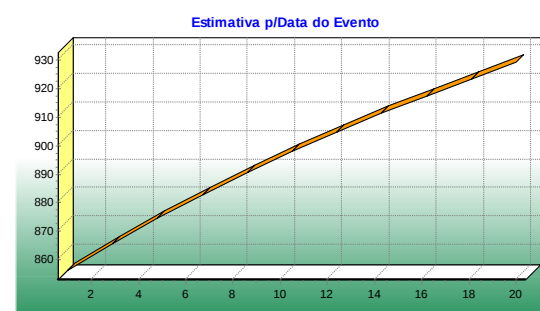
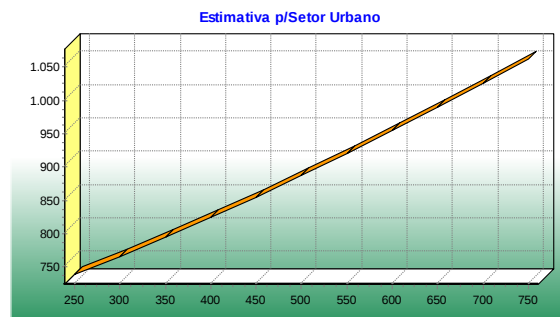
Redes Neurais Artificiais

W1	...									
-0,112	0,212	0,402	-0,218	0,319	-0,165	0,323	-0,031	-0,569	1,025	...0,747
-0,210	-0,371	0,611	0,072	-0,322	-0,825	-0,049	0,108	-0,444	0,036	...1,032
-0,499	0,028	0,222	-0,346	-0,071	-0,857	0,274	0,387	0,063	-0,104	0,399
-0,010	0,113	-0,356	0,089	0,469	0,175	-0,088	-0,132	0,089	0,203	...0,198
-0,027	0,027	-0,167	-1,077	0,342	-1,065	0,233	-0,210	0,592	0,580	...0,236
0,719	-0,159	-0,345	0,134	-0,811	-0,131	-0,289	-0,211	0,571	0,155	...0,571

w2						
0,956	-0,914	0,767	-0,596	-1,071	0,880	-0,023

Redes Neurais Artificiais

- Coeficiente correlação 0,8960
- Tendência normalidade
- Coerência hipóteses formuladas



Redes Neurais Artificiais

- Modelo de preços hedônicos
 - Valorização período (20 meses): 8,6%
 - Localização: 40% valor imóvel
- Grande variabilidade preços
 - Tentativas e erro

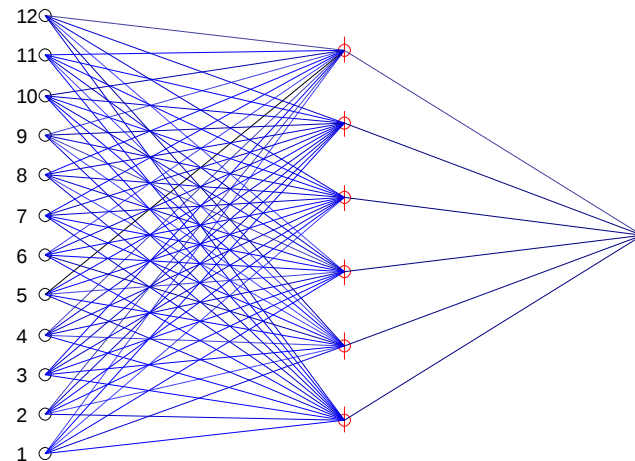
Redes Neurais Artificiais (Espacial)

$$P = f(L, E, T, w, WP) + \varepsilon$$

- onde f é indicativo da forma funcional, w os pesos a serem ajustados durante a fase de treinamento da rede, W a matriz de proximidade espacial e ε são os resíduos ou erros aleatórios. Serão utilizados para a modelagem os mesmos dados, porém sem transformações, do modelo MCR, incluindo a variável de defasagem espacial.

Redes Neurais Artificiais (Espacial)

Arquitetura da Rede 5



- Algoritmo de treinamento *Levenberg Marquadt*
- Estrutura rede: empírica (6 neurônios camada escondida)
- Função de transferência (tangente hiperbólica camada escondida, linear na camada de saída)
- Épocas (de 15 a 50), com erro quadrático mínimo 0,00001.

Redes Neurais Artificiais (Espacial)

TABELA 4.8 – Resultados da estimação com as RNA's

Variáveis	Valor Médio	Variação (%)
Setor Urbano	484,3142	3,21
Nível/Elevador	8,0575	3,39
Data do Evento	5,3407	1,36
Total Vagas	115,0442	3,79
Área Coberta	80,5880	-6,97
Área Descoberta	5,4513	4,23
Dormitórios	2,7566	1,96
Equipamentos	8,4735	2,37
Unidades p/ Acesso	11,4513	-1,41
Padrão Acabamento	250,7080	3,78
Estado Conservação	473,8938	2,38
WP	904,8713	2,34
Preço Unitário Estimado	910,36	

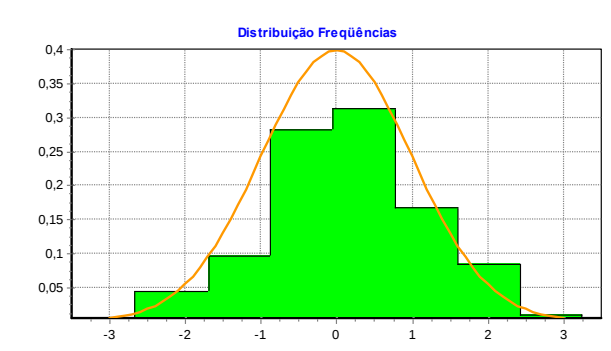
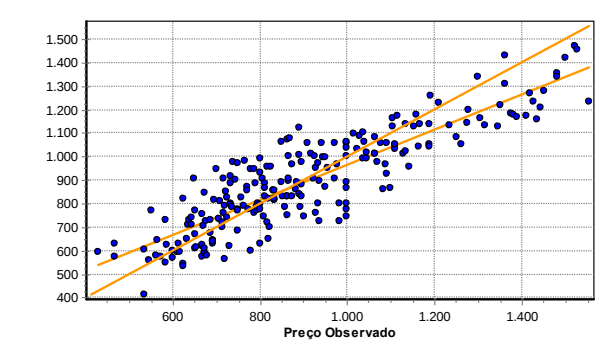
Redes Neurais Artificiais (Espacial)

W1												
-0,203	-0,360	-0,495	-0,716	-0,313	-0,040	-0,538	-0,235	0,554	-0,721	-0,032	0,449	0,107
-0,257	0,241	-0,251	-1,287	-0,345	-1,168	0,475	0,001	0,458	0,550	0,081	0,283	-0,649
0,272	0,472	0,099	-0,506	-0,296	-0,233	0,456	0,123	-0,298	0,673	-0,247	0,394	-0,143
-0,180	-0,025	0,184	-0,171	0,592	-0,303	-0,011	-0,110	0,233	-0,780	-0,538	-0,156	1,067
0,099	-0,303	-0,281	-0,116	0,250	0,340	-0,542	-0,170	-0,409	0,080	-0,060	0,202	0,676
-0,203	-0,360	-0,495	-0,716	-0,313	-0,040	-0,538	-0,235	0,554	-0,721	-0,032	0,449	0,107

w2					
0,903	-1,300	0,637	-0,809	-0,809	0,575

Redes Neurais Artificiais (Espacial)

- Coeficiente correlação 0,8780
- Tendência normalidade
- Coerência hipóteses formuladas



Redes Neurais Artificiais (Espacial)

- Modelo de preços hedônicos
 - Valorização período (20 meses): 11,36%
 - Localização: 43% valor imóvel
- Grande variabilidade valores estimados
 - Tentativas e erro

Validação dos Resultados

- Amostra inicial: 255 dados mercado
- Dados efetivamente utilizados: 231(dados discrepantes e outliers)
- Amostra subdividida randomicamente em 5 amostras
 - Cada amostra com dois conjuntos de dados
 - Treinamento / Modelagem – 189 dados
 - Validação – 42 dados

Validação dos Resultados

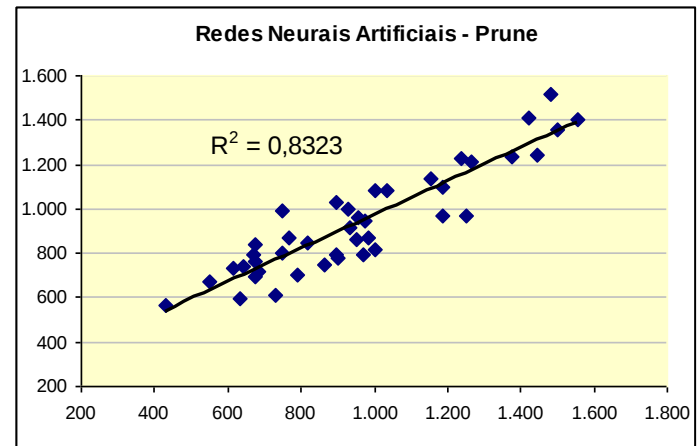
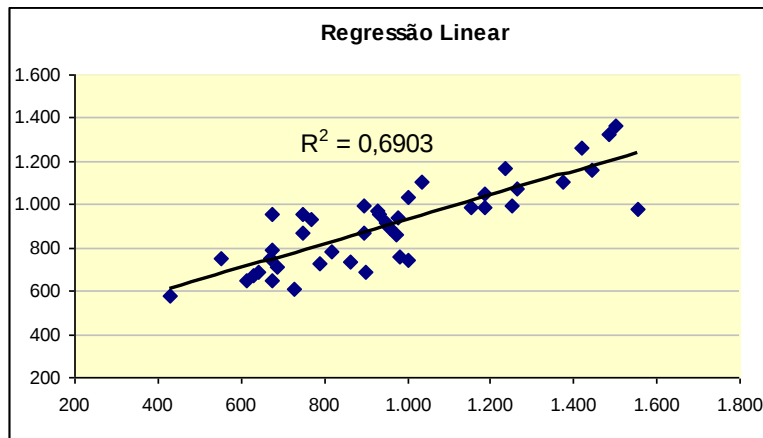
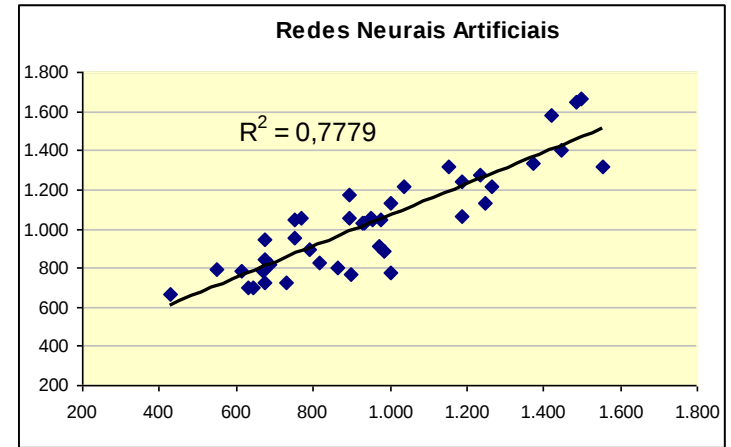
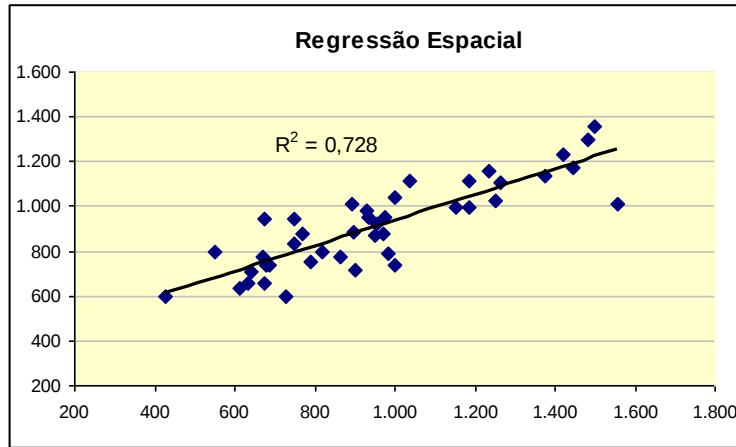
- Modelagens – 7 tratamentos diferenciados aos dados
 - 1º. regressão espacial, utilizando o modelo de defasagem espacial;
 - 2º. com o modelo regressão linear (MCR);
 - 3º. utilizando as Redes Neurais Artificiais, com as mesmas variáveis da regressão linear, sem aplicar as técnicas de poda;
 - 4º. utilizando as Redes Neurais Artificiais, com as mesmas variáveis da regressão linear, aplicando as técnicas de poda (utilizando o algoritmo OBS PRUNE);
 - 5º. utilizando as Redes Neurais Artificiais, com as mesmas variáveis da regressão espacial, aplicando as técnicas de poda (utilizando o algoritmo OBS PRUNE).
 - 6º. utilizando as Redes Neurais Artificiais, com as mesmas variáveis da regressão linear, aplicando Bagging.
 - 7º. utilizando as Redes Neurais Artificiais, com as mesmas variáveis da regressão espacial, aplicando Bagging.

Validação dos Resultados

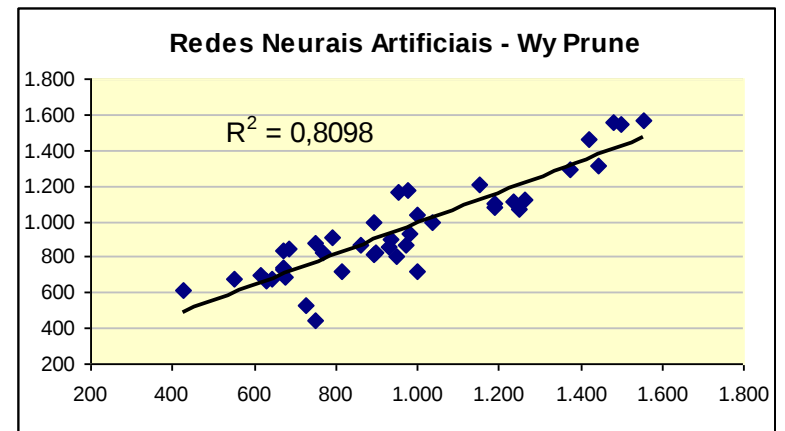
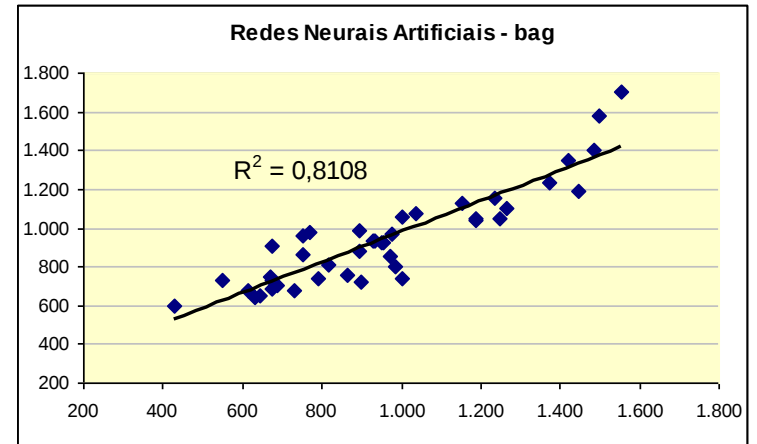
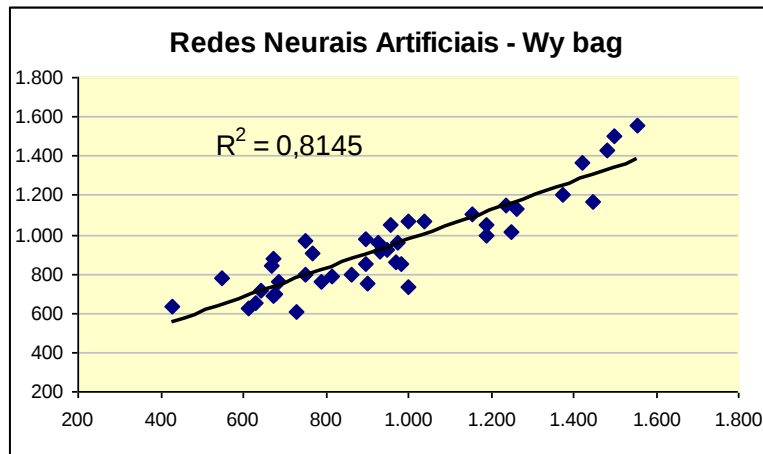
TABELA 4.11 – Resultados do processo de validação

	Resíduo Quadrático Médio para a validação				
Regressão Espacial	16.578	24.878	22.580	16.731	17.303
Regressão Linear	15.713	28.099	26.370	19.764	19.371
RNA	13.565	23.532	20.343	15.488	18.475
RNA Prune	13.394	15.622	17.173	15.627	14.215
RNA Wy Prune	15.785	15.662	17.078	17.052	15.157
RNA bagging		15.424			
RNA Wy bagging		15.771			

Validação dos Resultados



Validação dos Resultados



Validação dos Resultados

- Melhor resultado com as Redes Neurais Artificiais
- Modelo incluindo a variável Wy com a técnica de bagging
 - Consistência hipóteses formuladas
 - Menor variabilidade respostas
- Modelo de preços hedônicos
 - Valorização período (20 meses): 16,77%
 - (IPEAD-FACE UFMG no mesmo período: 17,73%)
 - Localização: 43% valor imóvel
- Modelo de preços hedônicos
 - Mais coerente mercado imobiliário

Conclusões e recomendações

- Conclusões
 - Importância introdução novos tratamentos aos dados de mercado
 - Avaliações em massa – evitar uso da regressão linear, na descrição do mercado imobiliário e nas estimativas de valores
 - Dificuldades na apresentação do modelo neural para fins de tributação
- Metodologias mais indicadas
 - Redes Neurais Artificiais (com especialista)
 - Regressão Espacial

Conclusões e recomendações

- Recomendações
 - Cuidados na elaboração matriz de pesos
 - Estudos para a construção de variáveis qualitativas utilizando algoritmos de agrupamento de classes
 - Inclusão da variável de defasagem espacial em modelos não lineares

Conclusões e recomendações

- Finalmente recomenda-se que, para elaboração de Avaliações em Massa caracterizadas pela não tendenciosidade, eficiência e consistência, seja utilizada a metodologia de *inferência espacial com a aplicação das Redes Neurais Artificiais e da Regressão Espacial*, evitando-se o uso da Regressão Linear.



XIX COBREAP | Foz do Iguaçu

INOVAÇÕES CIENTÍFICAS E TECNOLÓGICAS

AGOSTO 2017

Envelopamento de Dados

Análise da Envoltória de Dados – DEA

DEA: Ferramenta Benchmarking

- **Aplicado com sucesso para avaliar a eficiência de várias organizações e / ou processos.**
 - **Processo = Unidade para tomada de decisão (DMU)**
 - **A eficiência de um processo medida relativamente ao desempenho de outros processos no conjunto**
- **Considera processo como uma caixa preta e analisa as relações entre suas entradas e saídas**

processo de avaliação da empresa em relação à concorrência, por meio do qual incorpora os melhores desempenhos de outras firmas e/ou aperfeiçoa os seus próprios métodos.

Processo como Caixa Preta

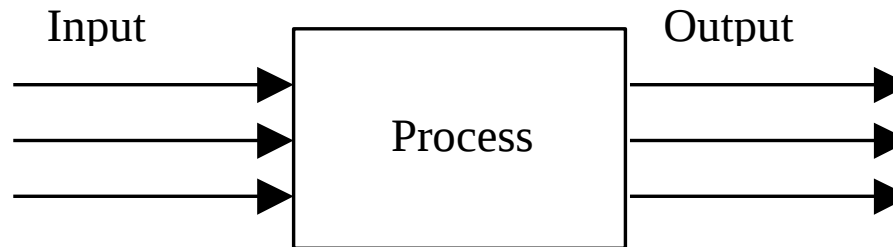


Figura 1 – Modelo de Caixa Preta como processo

$$\text{Eficiência} = \frac{\text{Output}}{\text{Input}}$$

- No entanto, com múltiplas entradas e saídas , torna-se mais difícil de avaliar a eficiência do processo.

Calculando Eficiência

Process	Custo trab. (\$/semana)	Saída (trab./semana)	Eficiência (trabalhos/\$)
A	2.000	1.500	0.750
B	1.500	1.100	0.733

- **Claramente , o processo A é mais eficiente do que o processo B , porém ...**

Processo	Área Escrit. (m ²)	Saída (trab./semana)	Eficiência (trab./m ²)
A	10,000	1,500	0.15
B	6,900	1,100	0.16

- **Uma nova avaliação com base no espaço de escritório mostra que o processo B é mais eficiente do processo de A, então ...**

Calculando Eficiência

- DEA oferece uma variedade de modelos que utilizam várias entradas e saídas para comparar a eficácia de dois ou mais processos.
- O modelo baseia-se na relação de definição de eficiência, mostrada a seguir:

$$\text{Eficiência} = \frac{\text{Soma ponderada das saídas}}{\text{Soma Ponderada das Entradas}}$$

Análise Gráfica

- Suponhamos que temos os seguintes dados de entrada e de saída:

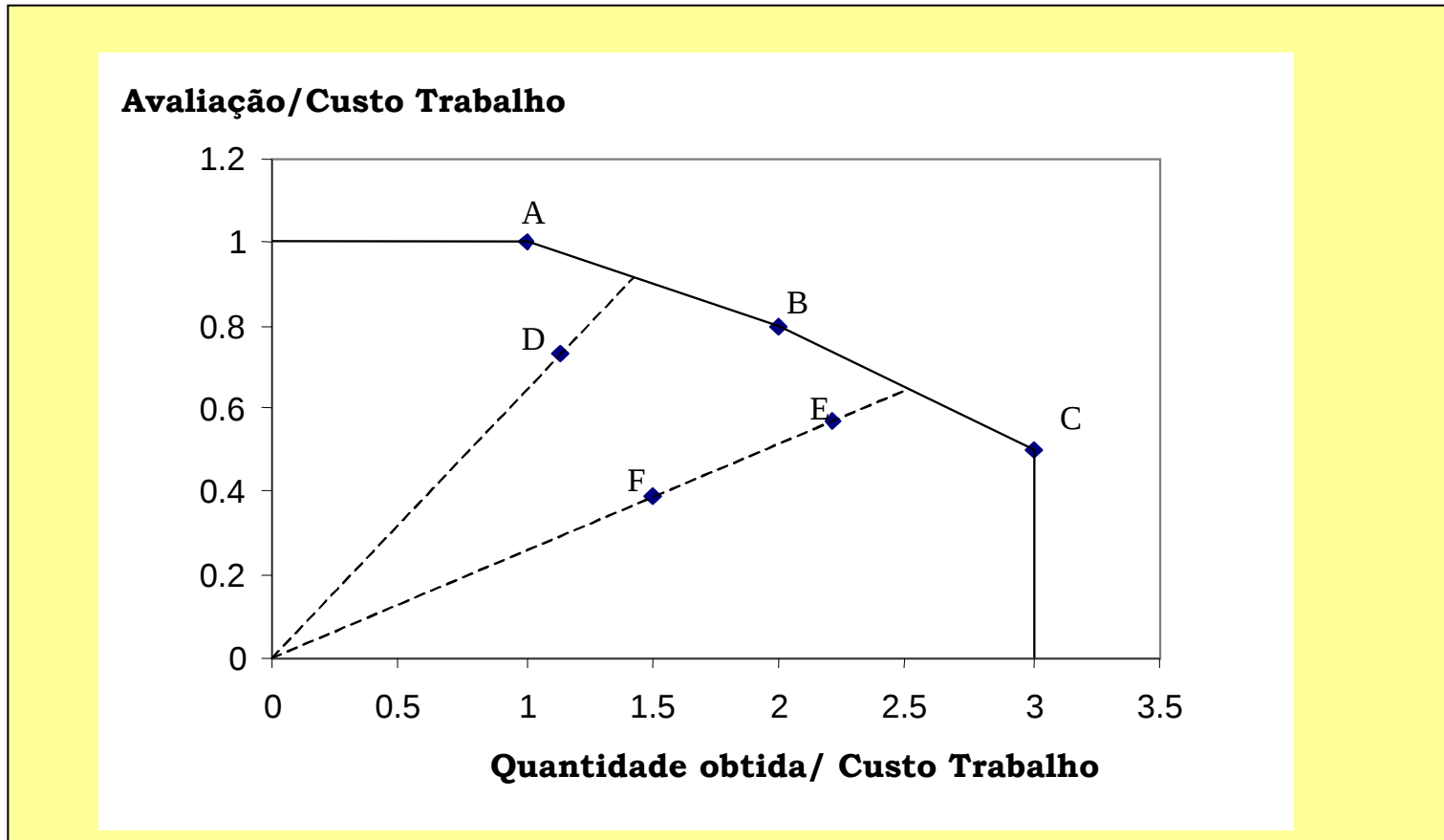
Processo	Custo Trabalho	Quant. Produzida	Avaliação Cliente
A	10	10	10
B	15	30	12
C	12	36	6
D	22	25	16
E	14	31	8
F	18	27	7

- Podemos rotular os raios de eficiência x e y:

Processo	Quant. Prod.	Avaliação
	$\frac{\text{Trabalho}}{\text{Trabalho}}$ (x)	$\frac{\text{Trabalho}}{\text{Trabalho}}$ (y)
A	1	1
B	2	0.8
C	3	0.5
D	1.136	0.727
E	2.214	0.571
F	1.5	0.389

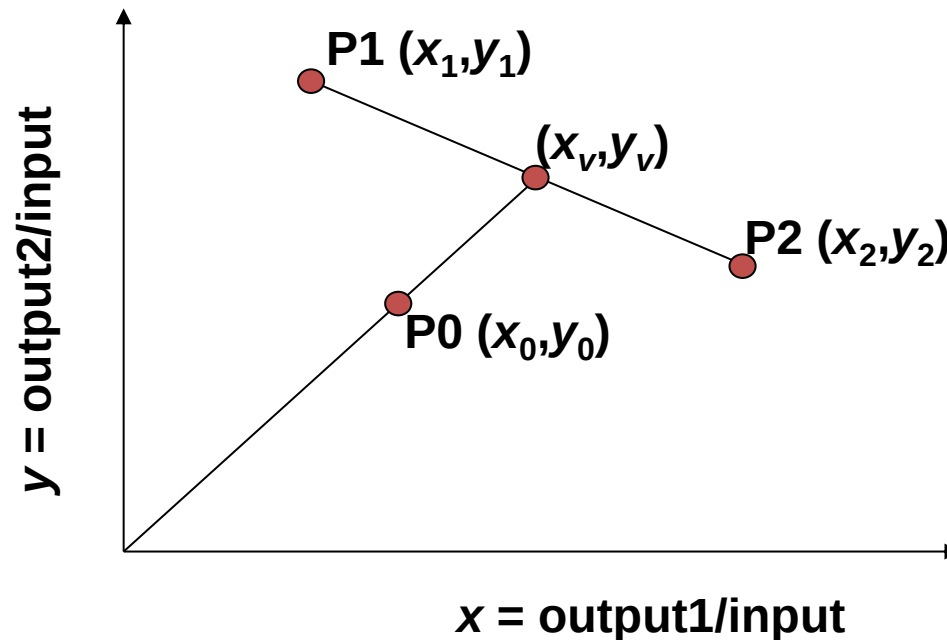
Análise Gráfica

- Então , marcamos a posição relativa de cada processo:



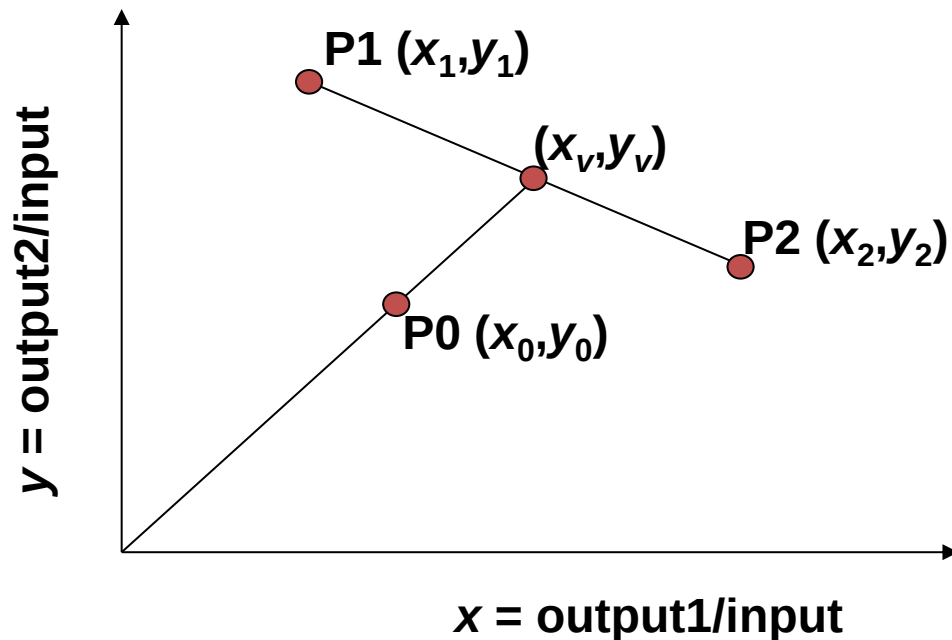
Calculo da Eficiência

- **Processos relativamente eficientes são aqueles situados na fronteira da eficiência:**
 - **Considerando uma eficiência de 100% .**
 - **E qual a ineficiência dos outros processos?**



Calculo da Eficiência

- P1 and P2 são relativamente eficientes → P0's *Grupo de Pares*.



Definindo a e b :

$$a = \frac{y_2 - y_1}{x_2 - x_1} \quad b = \frac{x_2 y_1 - x_1 y_2}{x_2 - x_1}$$

Então, obtemos a eficiência virtual do processo correspondendo a x_v e y_v :

$$x_v = \frac{b}{\frac{y_0}{x_0} - a} \quad y_v = a x_v + b$$

A eficiência do processo P0 é:

$$E_0 = \sqrt{\frac{x_0^2 + y_0^2}{x_v^2 + y_v^2}}$$

Programação Linear

- O modelo mede a proporção da eficiência de um processo por comparação com um processo hipotético que é uma combinação linear ponderada dos outros processos .
- Processos individuais poderá avaliar as entradas e saídas de maneira diferente.
- Portanto, cada processo é autorizado a adotar um conjunto de pesos para mostrá-lo sob a luz mais favorável

- Formulada como uma sequência de programas lineares (um para cada processo) para:
 - Maximizar a eficiência de um processo
 - Sujeito a eficiência de todos os processos $\leq 100\%$

Anexo D – EDO / DEA

- **Introdução**
- **D.1.1** A análise envoltória de dados é uma técnica consagrada para a avaliação de produtividade e eficiência, que se baseia em modelagem econométrica para a estimação de uma função de produção formada pelas unidades analisadas mais eficientes (unidades *benchmarking*).
- **D.1.1.1** A partir dos dados coletados, define-se o espaço viável de produção, delimitado pela envoltória representativa dos melhores resultados, orientada segundo a minimização dos insumos ou a maximização dos produtos. A eficiência de cada uma das unidades observadas é determinada pela distância normalizada à envoltória, através da solução de problemas de programação linear (PPL).

Anexo D – EDO / DEA

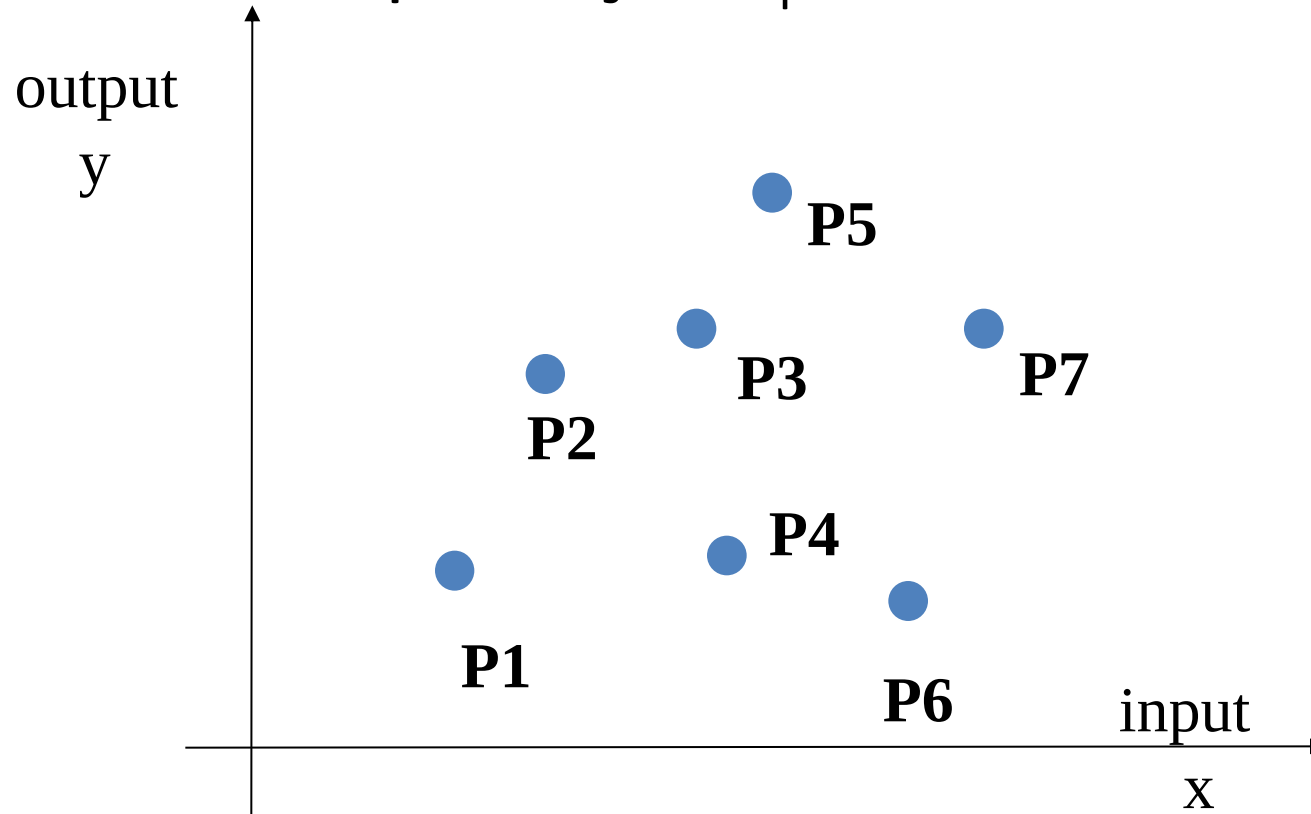
- **D.1.1.2** A análise envoltória de dados sob dupla ótica – EDO-DEA – explica as variações observadas em uma ou mais variáveis de interesse (variáveis dependentes), utilizando outras variáveis explicativas do fenômeno (variáveis independentes).

Anexo D – EDO / DEA

- **D.1.1.3** No caso de utilização da técnica para avaliações imobiliárias, na ótica do vendedor, os insumos são representados pelas características relevantes do imóvel e o produto, pelo seu preço. Na ótica do comprador, o insumo é o preço do imóvel e os produtos, suas características relevantes. Através dos PPL, são definidos os hiperplanos convexos que correspondem à envoltória da ótica do vendedor, onde ocorrem os maiores preços, e os que correspondem à envoltória da ótica do comprador, onde ocorrem os menores preços.

Anexo D – EDO / DEA

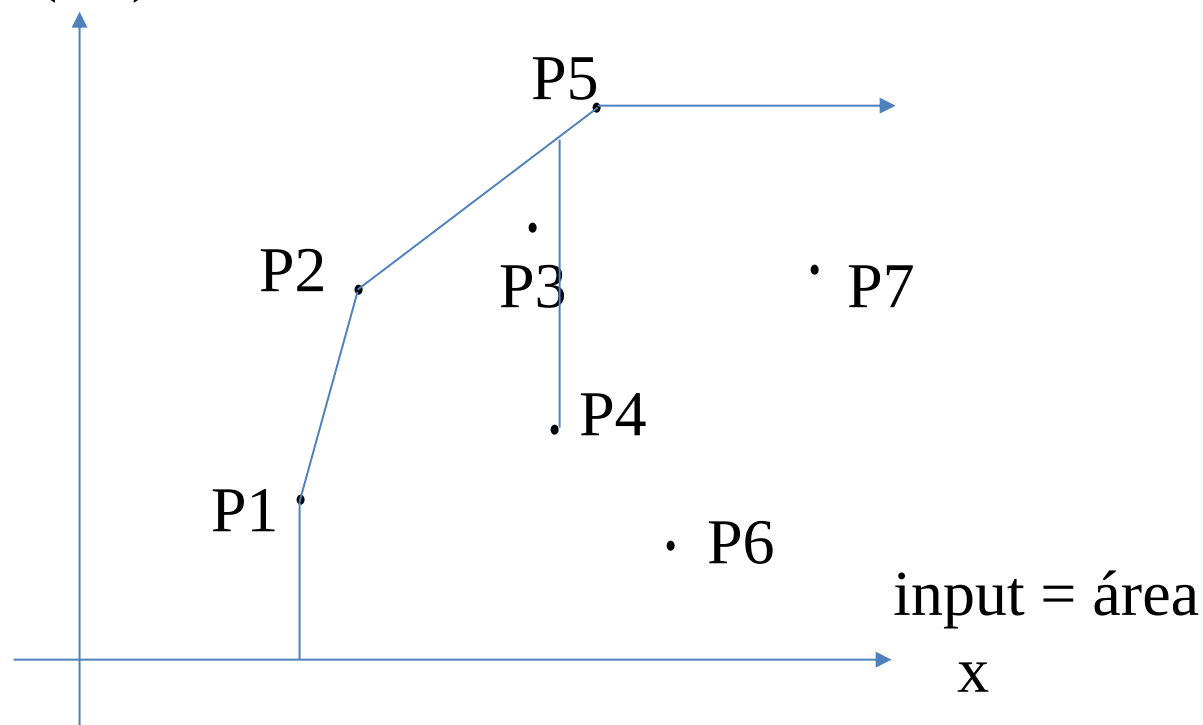
- Unidades de produção P_i



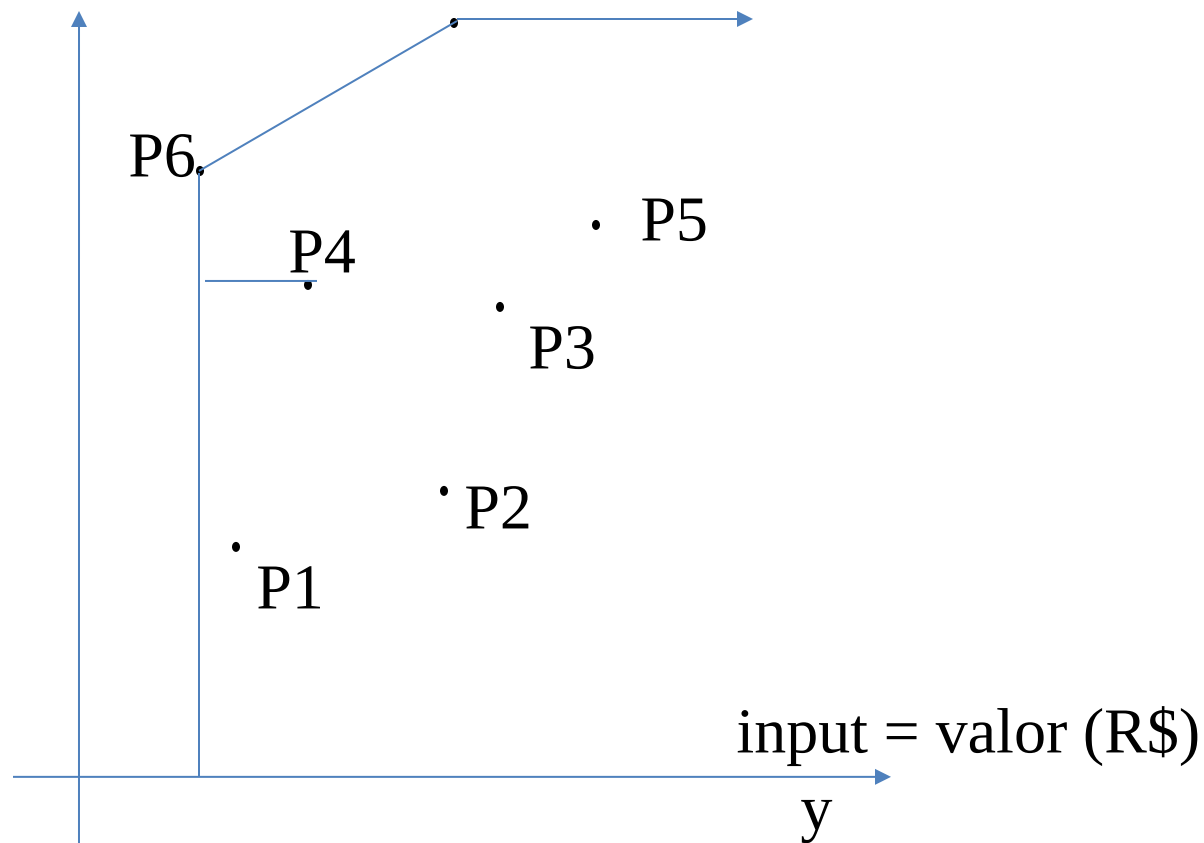
Anexo D – EDO / DEA

output = valor (R\$)

y

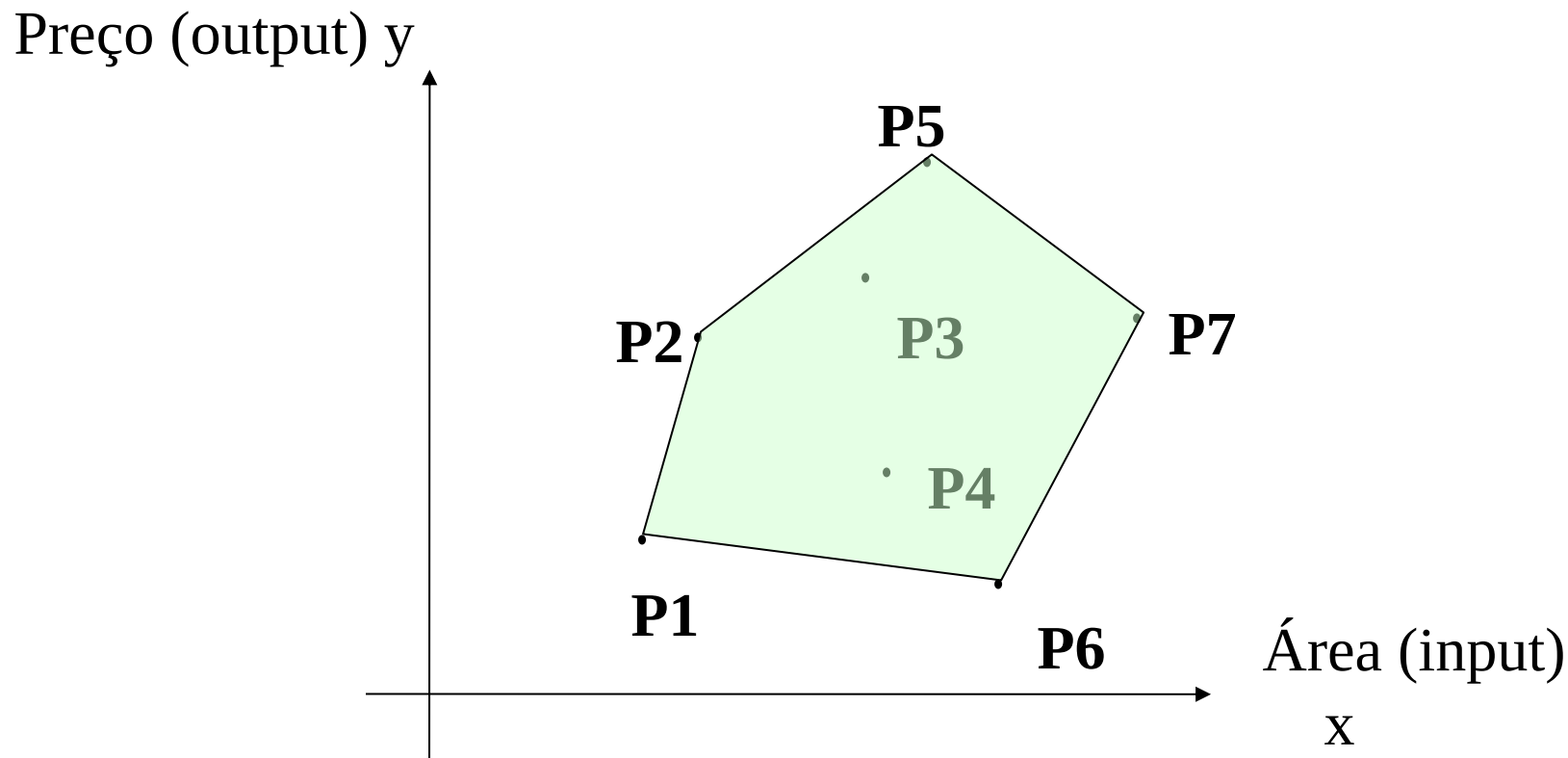


output = área
x

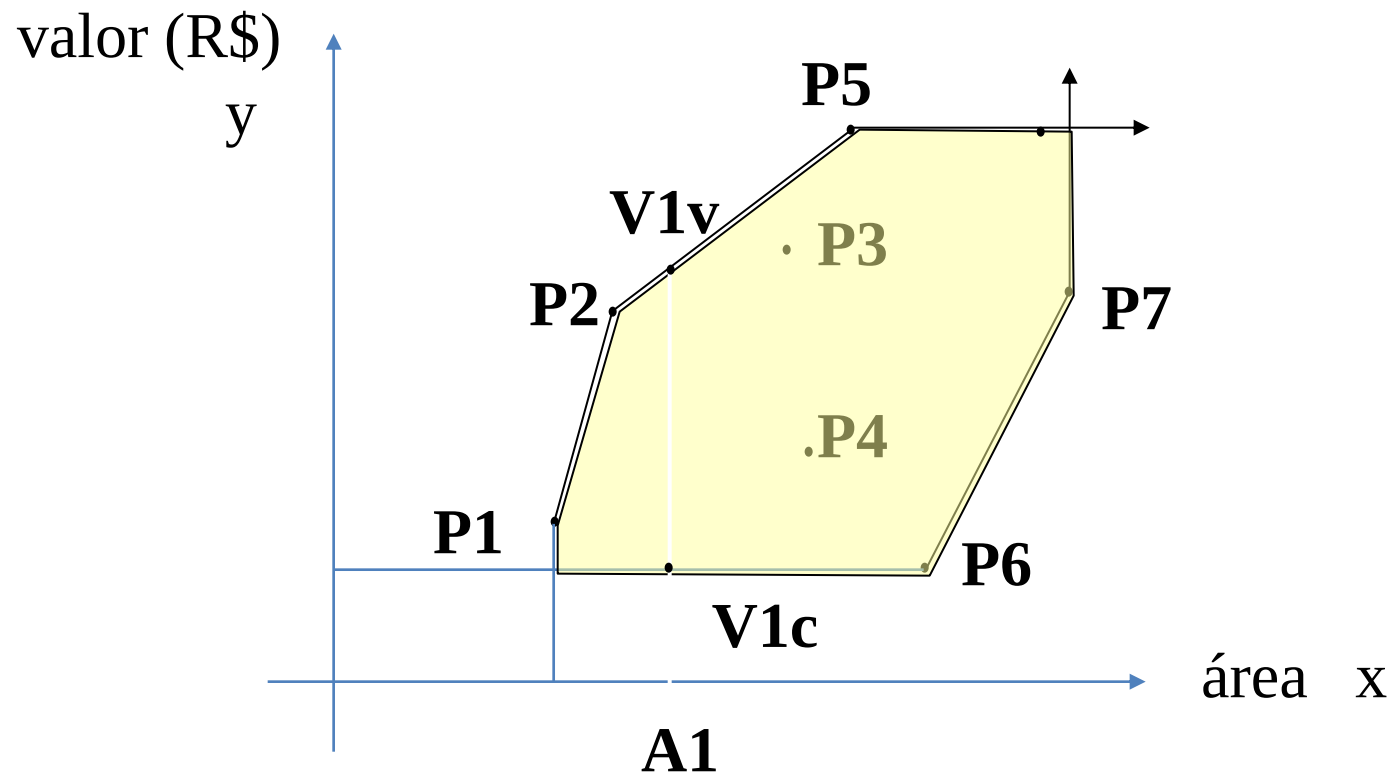


Anexo D – EDO / DEA

Convexidade



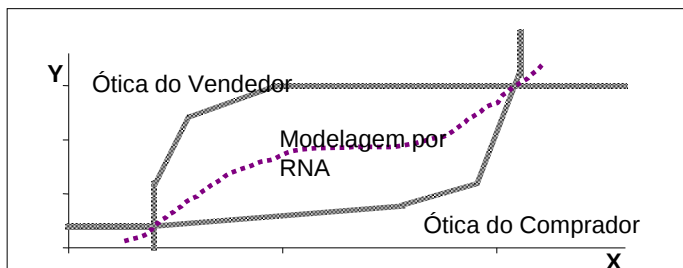
Anexo D – EDO / DEA



Anexo D – EDO / DEA

- Aplicação prática com uso do SisDEA
 - A utilização dessa ferramenta pode ser sintetizada da seguinte forma:
 - os bens que fazem parte da amostra colhida no mercado (unidades habitacionais – hotéis) são interpretados no ambiente da metodologia DEA, como as “unidades de produção”;
 - para o vendedor da unidade habitacional, ou seja, o hotel ou resort, os “insumos” (*inputs*) da sua “unidade de produção” são os atributos (características intrínsecas e extrínsecas) relevantes para a sua valorização, como localização, padrão, equipamentos disponíveis, sistema de alimentação & bebidas, etc.; o “produto” (*output*) da sua “unidade” é o preço, que pode ser expresso pelo preço da diária;
 - para o inquilino ou ocupante do imóvel, inversamente, o *input* é o preço e os *outputs* são os atributos (localização, padrão, equipamentos disponíveis, etc.);

Anexo D – EDO / DEA



Hotel	Valor Estimado	Tipo Apartamento: Duplo
A	R\$ 748,51	(superior / all inclusive)
B	R\$ 474,42	(superior / CM)
C	R\$ 697,38	(superior / MP)

Anexo D – EDO / DEA

- **D.1.2** O emprego da análise envoltória de dados sob dupla ótica (EDO-DEA) é especialmente útil para a realização de avaliações quando o tamanho da amostra de dados de mercado aproxima-se da própria população de eventos, para determinado tipo de imóvel, localização e período de tempo.
- O procedimento também é útil nos casos de análise de viabilidade e eficiência de empreendimentos, velocidade de vendas, análise de custos de construção de empreendimentos, avaliação em massa, entre outros.

Anexo D – EDO / DEA

- **Pressupostos básicos**
- **D.2.1** As variáveis independentes devem ter correlação positiva com a variável dependente. Caso isso não ocorra, devem ser realizadas transformações das variáveis independentes, de forma a alcançar esse pressuposto básico.
- **D.2.2** O número máximo de variáveis independentes (k) efetivamente utilizadas no modelo deve obedecer ao seguinte critério:
 - $k < n / 3$

Anexo D – EDO / DEA

- **Recomendações**
- **D.3.1** No caso de utilização de EDO/DEA em amostra com tamanho muito inferior ao da população, recomenda-se verificar se a estimativa pontual encontra-se dentro do intervalo de confiança a 80 % calculado com o uso de modelo de regressão linear.
- **D.3.2** Recomenda-se que, no caso de utilização de variáveis expressas por dicotomias ou códigos alocados, haja pelo menos três dados para a mesma característica de cada variável.

Anexo D – EDO / DEA

- **D.3.3** A importância de cada uma das variáveis pode ser identificada pelo exame dos respectivos pesos, calculados pelo Problema de Programação Linear (PPL) segundo o modelo dos multiplicadores, em cada ótica.
- **D.3.4** Possíveis dados de mercado atípicos devem ser investigados e sua retirada do modelo fica condicionada à apresentação de justificativas. Deve-se observar que um dado atípico pode tornar uma envoltória menos aderente à massa de dados. Nesse sentido, recomenda-se que pelo menos 20 % dos dados de mercado pertençam a cada uma das envoltórias.

Anexo D – EDO / DEA

- **D.3.5** Uma forte dependência linear entre duas variáveis independentes pode provocar degenerações no modelo EDO/DEA e limita a sua utilização.
- **D.3.6** Para verificação da colinearidade, recomenda-se a análise da matriz de correlações, que espelha as dependências lineares de primeira ordem entre as variáveis independentes, com atenção especial para resultados superiores a 0,80.

Anexo D – EDO / DEA

- Para tratar dados na presença de colinearidade, é recomendável que sejam tomadas medidas corretivas, como a ampliação da amostra, a eliminação da variável redundante ou o uso de componentes principais.