A ANÁLISE DE CAMINHOS DA REGRESSÃO LINEAR MÚLTIPLA COMO SUBSÍDIO AO TRATAMENTO DE HOMOGENEIZAÇÃO POR FATORES

LIMA, GILSON PEREIRA DE ANDRADE
Eng. Civil, *M.Sc.*Eng$^{\underline{ia}}$.de Produção
CREA nº 27.600-D/RJ
IEL-RJ nº 1.298
IBAPE-SP nº 812
UERJ/FACULDADE DE ENGENHARIA
Rua Augusto do Amaral Peixoto nº 213 aptº 504, Teresópolis, RJ, CEP 25961-167
Tel. 0-xx-21-26429361, Fax: 0-xx-21-25877584 e-mail: gpal@uol.com.br

CURRICULUM VITAE

GILSON PEREIRA DE ANDRADE LIMA

51 anos, Engenheiro Civil graduado pela Pontifícia Universidade Católica do Rio de Janeiro (1974) e Mestre em Ciências em Engenharia de Produção na área de Projetos Industriais pela COPPE/UFRJ, tendo defendido Tese de Mestrado em maio/1992 com o título "Avaliação de Bens sob a Ótica da Análise de Investimentos em Condições de Risco", e atualmente cursando o Doutorado em Planejamento Energético no Programa de Planejamento Energético da COPPE/UFRJ.

Profissional com 27 anos no magistério do nível superior em Faculdades de Engenharia e um total de 29 anos de atuação como engenheiro, sendo 13 anos nas áreas de Projetos, Fiscalização e Contratação de Serviços de Engenharia e 16 anos na área de Avaliações Técnicas de Bens.

DESTAQUES NA ATUAÇÃO PROFISSIONAL

Consultor Técnico do CEFEN - Centro de Estudos da Faculdade de Engenharia da UERJ (2002/atual).

Professor da Faculdade de Engenharia da Universidade do Estado do Rio de Janeiro (1979/atual), tendo ministrado inicialmente as disciplinas Concreto Armado I e II, Complementos de Concreto Armado e Concreto Protendido do curso de graduação em Engenharia Civil, posteriormente, de 1989 a 1998, a disciplina Tópicos Especiais em Construção Civil I do curso de graduação em Engenharia Civil, cujo tema é a Engenharia de Avaliações e, desde janeiro/1999, a disciplina Engenharia de Custos, do curso de graduação em Engenharia Civil.

Ex-empregado da PETROBRÁS S.A. (1975/2001), tendo atuado em fiscalização e execução de projetos de estruturas marítimas, fiscalização e execução de projetos de estruturas e fundações em concreto armado e aço (1975/1986), tendo atuado, a partir de 1987, no Setor de Engenharia de Perícias e Avaliações (SEPAV) da unidade de Engenharia, na elaboração de avaliação técnica de bens (setor imobiliário urbano/rural e setor industrial), tendo encerrado a carreira na função de Consultor Técnico de Avaliação de Mercado e Econômicas da Unidade de Engenharia da Petrobras.

Gilson Pereira de Andrade Lima
Engª Civil, M. Sc. Engenharia de Produção
CREA/RJ-27.600-D
Membro do Instituto de Engenharia Legal - IEL-RJ, registro nº 1.298
Membro do Instituto Brasileiro de Avaliações e Perícias de Engenharia de São Paulo - IBAPE-SP, registro nº 812
Rua Augusto do Amaral Peixoto nº 213 aptº 504, Teresópolis, RJ, CEP 25961-167
Tel. 0-xx-21-26429361, Fax: 0-xx-21-25877584 e-mail: gpal@uol.com.br

A ANÁLISE DE CAMINHOS DA REGRESSÃO LINEAR MÚLTIPLA COMO SUBSÍDIO AO TRATAMENTO DE HOMOGENEIZAÇÃO POR FATORES

Resumo.
Este trabalho tem por finalidade apresentar a análise de caminhos da regressão linear múltipla e as suas conseqüências na utilização dos fatores de homogeneização fundamentados originários de modelos de regressão, alertando para o fato de que se na homogeneização por fatores forem utilizados menos variáveis independentes que no modelo de regressão que originou os fatores, os fatores de homogeneização devem ser recalculados levando em conta os efeitos indiretos das variáveis não consideradas e que, mesmo nestes casos, variáveis importantes não podem ser omitidas.

Abstract.
This paper objective to explain the multiple regression path analysis and theirs consequences in using the fundamental factors of homogenization based on multiple regression models, and alert to the fact that in homogenization model by factors using fewer independents variables as the multiple regression that was the origin of the factor, these factors need to be recalculated, including the indirect effects of the non considering variables and, even so, significant variables can't to be neglected.

Palavras-chave: Homogeneização, Fatores, Análise de Caminhos, Regressão Linear

A ANÁLISE DE CAMINHOS DA REGRESSÃO LINEAR MÚLTIPLA COMO SUBSÍDIO AO TRATAMENTO DE HOMOGENEIZAÇÃO POR FATORES

1 - INTRODUÇÃO

Na avaliação de bens, quando é utilizado o método comparativo direto de dados de mercado[1] com tratamento por modelos de homogeneização por fatores, e estes fatores são fundamentados[2], derivados de estudos de mercado por modelos de regressão linear múltipla, o modelo adotado para o tratamento de dados pode ou não ser adequado, se tornando menos adequado quando no tratamento de homogeneização por fatores forem utilizados menos variáveis que no modelo de regressão que originou os fatores.

A finalidade deste trabalho é apresentar e aplicar a técnica de análise de caminhos da regressão linear múltipla, que vai nos esclarecer que, nos casos em que no tratamento de homogeneização por fatores forem utilizados menos variáveis que no modelo de regressão que originou os fatores, os fatores de homogeneização devem ser recalculados, incorporando o viés das variáveis não consideradas, para que o modelo fique mais adequado, e será cada vez mais adequado à medida que não se omitam variáveis importantes.

A comprovação disto pode ser feita pela análise dos resíduos, seja na forma gráfica[3], seja pelo coeficiente de homogeneidade[4].

2. A ANÁLISE DE CAMINHOS DA REGRESSÃO LINEAR MÚLTIPLA.

A análise de caminhos é uma generalização da análise de efeitos diretos e indiretos das variáveis independentes, partindo da idéia de que os parâmetros da regressão são fatores de multiplicação.

2.1 OS PARÂMETROS DA REGRESSÃO COMO FATORES DE MULTIPLICAÇÃO

O parâmetro B_i de um modelo de regressão linear tem uma simples interpretação de que é a variação da variável dependente $\Delta\hat{Y}_i$ para uma variação unitária da variável independente X_i ($\Delta X_i = 1$), pois:

$$\hat{Y}_i = B_0 + B_1 X_1 + ... + B_i X_i + ... + B_k X_k \qquad \text{(eq.2.1)}$$

e para uma variação de ΔX_i em X_i teremos:

$$\hat{Y}_i + \Delta\hat{Y}_i = B_0 + B_1 X_1 + ... + B_i X_i + B_i \Delta X_i + ... + B_k X_k$$

---

[1] Aquele que identifica o valor de mercado do bem por meio de tratamento técnico dos atributos dos elementos comparáveis, constituintes da amostra. (fonte: NBR-14653 Avaliação de Bens – Parte 1: Procedimentos Gerais, ABNT, 2001).

[2] Ver Lima, Gilson Pereira de Andrade. Homogenheização Fundamentada - Uma Utopia? VIII Congresso Brasileiro de Engenharia de Avaliações e Perícias, Florianópolis, SC, 1995.

[3] Ver Lima, Gilson P.de A., Separando o Joio do Trigo: A Importância da Análise de Resíduos na Homogeneização por Fatores, Anais do 1º Seminário Internacional de Real State da América Latina da LARES – Latin American Real State Society, São Paulo, SP, 1999.

[4] Ver Lima, Gilson P.de A., O coeficiente de Homogeneidade na Homogeneização por Fatores, Anais do 2º Seminário Internacional de Real State da América Latina da LARES – Latin American Real State Society, São Paulo, SP, 2001.

e por diferença temos:

$\Delta\hat{Y}_i = B_i \, \Delta X_i$ (eq.2.2)

e quando $\Delta X_i = 1$, temos $\Delta\hat{Y}_i = B_i$

Como vimos o que ocorre quando somente uma variável independente é alterada, podemos estender a análise e concluir que a mudança total em $\hat{Y}$ é simplesmente a soma das mudanças individuais de cada variável independente:

$\Delta\hat{Y}_i = B_1 \, \Delta X_1 + \ldots + B_i \, \Delta X_i + \ldots + B_k \, \Delta X_k$ (eq.2.3)

2.2 O PARÂMETRO DO MODELO DE REGRESSÃO SIMPLES COMO EFEITO TOTAL

Os valores obtidos para o parâmetro da variável independente nos modelos de regressão linear simples são, normalmente, diferentes dos obtidos com a regressão múltipla ao acrescentar outras variáveis independentes ao modelo, isto porque no modelo de regressão simples o parâmetro mostra como a variável dependente é afetada pela variação da variável independente contemplada no modelo, conjuntamente com as variações das outras variáveis independentes associadas não contempladas no modelo.

Para provar de forma geral que o parâmetro do modelo de regressão simples fornece o efeito total, soma do efeito direto e do indireto, podemos partir das equações que resultam no cálculo dos parâmetros do modelo de regressão múltipla, com duas variáveis independentes X_1 e X_2.

A regressão estimada da amostra com “n” registros é da forma:

$Y_i = B_0 + B_1 X_{1i} + B_2 X_{2i} + R_i$,i = 1,n (eq.2.4)

onde "B_0", "B_1" e "B_2" são os parâmetros da regressão e $R_i = Y_i - \hat{Y}_i$ é o resíduo entre a o valor efetivo Y_i e o valor estimado $\hat{Y}_i$

Substituindo "R_i" na equação resulta a regressão linear da amostra:

$\hat{Y}_i = B_0 + B_1 X_{1i} + B_2 X_{2i}$,i = 1,n (eq.2.5)

Ao ajustar a regressão amostral, um objetivo razoável é manter os resíduos "R_i" tão pequenos quanto possível , o que pode ser feito pelo método dos mínimos quadrados. que consiste em minimizar o somatório do quadrado dos resíduos "R_i", ou seja, minimizar a soma "SQR" dada por:

$SQR = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2$ (eq.2.6)

ou

$$SQR = \sum_{i=1}^{n}\left(Y_i - B_0 - B_1 . X_{1i} - B_2 . X_{2i}\right)^2 \qquad \text{(eq.2.7)}$$

Diferenciando "SQR" com relação aos parâmetros "B_0" "B_1" e "B_2" e igualando a zero obteremos três equações em "B_0" "B_1 " e "B_2" que nos permitirão estimar este parâmetros de modo que "SQR" seja mínimo, dadas por:

$$\sum_{i=1}^{n} x_{1i} . y_i = B_1 . \sum_{i=1}^{n} x_{1i}^{\;2} + B_2 \sum_{i=1}^{n} x_{1i} . x_{2i} \qquad \text{(eq.2.8)}$$

$$\sum_{i=1}^{n} x_{2i} . y_i = B_1 . \sum_{i=1}^{n} x_{1i} x_{2i} + B_2 \sum_{i=1}^{n} x_{2i}^{\;2} \qquad \text{(eq.2.9)}$$

$$B_0 = \overline{Y} - B_1 . \overline{X}_1 - B_2 . \overline{X}_2 \qquad \text{(eq.2.10)}$$

onde representamos os desvios em relação às médias $(\overline{Y}, \overline{X}_1, \overline{X}_2)$ pelas notações abreviadas dadas por:

$$Y_i - \overline{Y} = y_i$$

$$X_{1i} - \overline{X_1} = x_{1i}$$

$$X_{2i} - \overline{X_2} = x_{2i}$$

Dividindo a equação (2.8) por $\sum_{i=1}^{n} x_{1i}^{\;2}$, resulta :

$$\frac{\sum_{i=1}^{n} x_{1i} . y_i}{\sum_{i=1}^{n} x_{1i}^{\;2}} = B_1 . + B_2 \frac{\sum_{i=1}^{n} x_{1i} . x_{2i}}{\sum_{i=1}^{n} x_{1i}^{\;2}} \qquad \text{(eq.2.11)}$$

Como:

$\dfrac{\sum_{i=1}^{n} x_{1i} . y_i}{\sum_{i=1}^{n} x_{1i}^{\;2}}$ = parâmetro B_1 da variável X_1 do modelo de regressão simples entre Y e X_1, ou seja, é o efeito total de X_1 em relação a Y obtido do modelo $\hat{Y}_i = B_0 + B_1 X_1$

$\dfrac{\sum_{i=1}^{n} x_{1i}.x_{2i}}{\sum_{i=1}^{n} x_{1i}^{2}}$ = parâmetro B da variável X_1 do modelo de regressão simples entre X_2 e X_1, ou seja, é o efeito da variável X_1 em X_2, obtido do modelo $\acute{X}_2 = B_0 + B\ X_1$

Então, a equação 2.11 pode ser re-escrita como:

Efeito total de $X_1 = B_1 + B_2 \ .\ B$ (eq.2.12)

O parâmetro B_1 é o efeito direto da variável X_1 em Y, mantida a variável X_2 constante, o parâmetro B_2 é o efeito direto da variável X_2 em Y, mantida a variável X_1 constante e o parâmetro B_2 .B pode ser interpretado como o efeito indireto da variável X_2 em Y.

Então, a equação 2.12 pode ser reescrita como:

Efeito total de X_1 = efeito direto de X_1 + efeito indireto de X_2 (eq.2.13)

A vantagem desta decomposição de efeitos é que com ela podemos identificar suas importâncias relativas.

Podemos também identificar o viés, ou seja, a tendenciosidade sistemática que irá ocorrer se erroneamente utilizarmos o parâmetro da regressão simples (efeito total) quando deveríamos utilizar o parâmetro da regressão múltipla (efeito direto).

O efeito indireto que nos desvia do alvo é o viés ou tendenciosidade, e a variável responsável por isto é a variável mascarada.

Resumindo, se quisermos o efeito da variação de uma variável, mantendo as demais constantes, temos que incluir estas variáveis no modelo de regressão múltipla. Se não, as variáveis não incluídas se tornarão mascaradas, contribuindo para o viés, ou desvio do alvo.

Quanto mais variáveis mascaradas omitimos num estudo observacional, mesmo não intencionalmente, mais risco de viés teremos, sendo o modelo mais arriscado o da regressão simples (que omite todas as demais variáveis).

Incluir o máximo possível de variáveis independentes no modelo reduz, ou até elimina, o viés, mas nem sempre isto é possível, seja pelo tamanho da amostra, seja pela colinearidade entre variáveis independentes.

Existe outra maneira de eliminar o viés, que é planejar o experimento ou coletar dados que apresentem tão pouco relacionamento quanto possível entre os regressores para tentar eliminar a correlação entre as variáveis independentes, uma vez que inexistindo a correlação, o parâmetro do modelo de regressão simples entre as variáveis independentes é nulo e o efeito indireto se anula. Mas nem sempre isto também é possível, principalmente na avaliação de imóveis, que é um estudo observacional.

2.2.1 EXEMPLO NUMÉRICO DO PARÂMETRO DO MODELO DE REGRESSÃO SIMPLES COMO EFEITO TOTAL

Para um exemplo numérico se resgata o apresentado em Lima[5] (1999).

[5] Lima, Gilson Pereira de Andrade. Separando o Joio do Trigo: A Importância da Análise de Resíduos na Homogeneização por Fatores, Anais do 1º Seminário Internacional de Real State da América Latina da LARES – Latin American Real State Society, São Paulo, SP, 1999.

Tratava-se de um exemplo onde para avaliar um lote, o profissional tinha coletado uma amostra com dez registros, todos de transações concretizadas recentemente envolvendo lotes com características semelhantes ao avaliando (em termos de topografia, infra-estrutura, acesso, etc.), apresentada na tabela 1:

REGISTRO	DISTÂNCIA (m)	ÁREA (m^2)	PREÇO TOTAL (R$)	PU (R$/$m^2$)
1	2.200,00	300,00	30.000,00	100,00
2	2.000,00	340,00	37.400,00	110,00
3	1.800,00	270,00	32.400,00	120,00
4	1.500,00	360,00	50.400,00	140,00
5	2.300,00	400,00	34.000,00	85,00
6	1.900,00	500,00	52.500,00	105,00
7	1.300,00	600,00	72.000,00	120,00
8	2.200,00	300,00	28.500,00	95,00
9	900,00	360,00	54.000,00	150,00
10	1.700,00	600,00	60.000,00	100,00

Tabela 1 – amostra com fatores de diferenciação distância à orla e área

Os únicos atributos de diferenciação levantados compreendem a localização do lote, através da distância do mesmo à orla marítima em metros e a área do lote.

Para contrastar os diferentes usos da regressão simples e múltipla foram desenvolvidos os modelos seguintes:

$$\widehat{PU} = 184{,}161 - 0{,}040259\ DIST \qquad \text{(eq.2.14)}$$

p-valor (bi-caudal) = 0,00076939

$$\widehat{PU} = 119{,}060 - 0{,}016279\ AREA \qquad \text{(eq.2.15)}$$

p-valor (bi-caudal) = 0,788399

$$\widehat{PU} = 222{,}789 - 0{,}046252\ DIST - 0{,}069381\ AREA \qquad \text{(eq.2.16)}$$

p-valor (bi-caudal) = 0,0000261338 0,0052483

$$\widehat{DIST} = 2242{,}691 - 1{,}148118\ AREA \qquad \text{(eq.2.17)}$$

p-valor (bi-caudal) = 0,3754816

$$\widehat{AREA} = 556{,}745 - 0{,}086374\ DIST \qquad \text{(eq.2.18)}$$

p-valor (bi-caudal) = 0,3754816

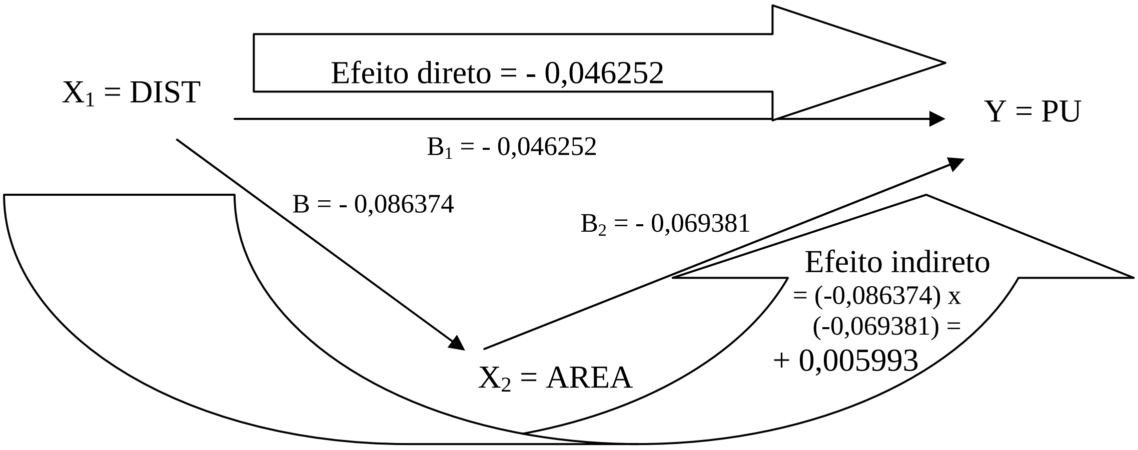
Para um lote com uma determinada área, a expectativa de variação do valor em função da distância à orla, dada pela equação 2.16, é de que para cada aumento de, por exemplo, 100 m na distância, o valor unitário caia cerca de R$ 4,63/m^2.

Para um lote com uma certa distância à orla, a expectativa de variação do valor em função da área, dada pela equação 2.16, é de que para cada aumento de, por exemplo, 100 m^2 na área, o valor unitário caia cerca de R$ 6,94/m^2.

Mas se a tendência estimada pela regressão simples entre as variáveis DIST e AREA (eq. 2.17 e eq. 2.18) de que à medida que a distância aumenta a área diminui e vice-versa persiste, o que se pode esperar para a expectativa de variação do valor em função da variação da distância à orla e da área? Neste caso a resposta teria que ser dada pela regressão linear simples, pois não estaremos querendo o efeito da variação de uma variável mantendo a outra variável inalterada, mas sim o efeito da variação de uma variável com a outra variando segundo a tendência de variação conjunta entre elas. Neste caso, pela equação 2.14, para cada aumento de, por exemplo, 100 m na distância, o valor unitário cai cerca de R$ 4,03/ m^2 (enquanto a área cai de 8,63 m^2, conforme a equação 2.18), ao invés da queda de R$ 4,63/m^2 obtida na regressão múltipla (efeito isolado da variação da distância à orla, mantida a área). Para cada aumento de, por exemplo, 100 m^2 na área, pela equação 2.15, o valor unitário cai cerca de R$ 1,63/m^2 m^2 (enquanto a distância cai de 14,81 m, conforme a equação 2.17), ao invés da queda de R$ 6,94/m^2 obtida na regressão múltipla (efeito isolado da variação da área, mantida a distância à orla).

Os valores obtidos com as regressões simples foram diferentes dos obtidos com a regressão múltipla porque no modelo de regressão simples o parâmetro mostra como o valor unitário é afetado pela variação da variável independente contemplada no modelo conjuntamente com as variações das outras variáveis independentes associadas não contempladas no modelo.

Para uma melhor compreensão podemos utilizar um diagrama, apresentado na figura 1, para representar o efeito direto e indireto da distância no preço unitário:



$$\hat{PU} = 222{,}789 - 0{,}046252\ DIST - 0{,}069381\ AREA$$

$$\hat{AREA} = 556{,}745 - 0{,}086374\ DIST$$

$$\hat{PU} = 184{,}161 - 0{,}040259\ DIST$$

Figura 1 – Diagrama de efeito direto e indireto da regressão

Como a distância à orla varia conjuntamente com a área, a expectativa de variação do preço unitário pode ser obtida com a regressão simples do preço unitário em relação à distância à orla, mas isto também pode ser obtido pela soma do efeito direto da variação unitária da distância à orla (mantida a área constante) com o efeito indireto da variação unitária da distância à orla via a variação correspondente da variável área.

O efeito direto da variação unitária da distância à orla (mantida a área) é representado pelo regressor B_1 da variável distância no modelo da regressão múltipla (B_1 = - 0,046252).

O efeito indireto da variação unitária da distância à orla via a variação correspondente da variável área é representado pela composição do regressor B da regressão simples da área em função da distância (B = - 0,086374) e o regressor B_2 da variável área no modelo da regressão múltipla (B_2 = - 0,069381), resultando B x B_2 = (- 0,086374) x (- 0,069381) = + 0,005993.

Então, somando os dois efeitos, resulta: -0,046252 + 0,005993 = - 0,040259

Este resultado é coincidente com o regressor B_1 da regressão simples do preço unitário em função da distância (B_1 = - 0,040259) obtido na equação 2.14.

No caso deste exemplo numérico que estamos desenvolvendo, se quisermos saber o efeito no preço unitário da variação da distância à orla, mantidas as demais variáveis (no caso, a área dos lotes) constantes, esta estimativa é dada pelo parâmetro do modelo de regressão múltipla B_1 = - 0,046252, que resulta da regressão do preço unitário (PU) com ambas as variáveis independentes, distância à orla (DIST) e área dos lotes (AREA).

Mas se a variável área dos lotes não tivesse sido coletada na amostra, ou mesmo, se coletada, não tivesse sido cogitada a sua influência no preço unitário do lote, a regressão simples do preço unitário com a distância à orla (DIST) poderia ter sido modelada, e captaria tanto o efeito direto desejado (B_1 = - 0,046252) quanto o efeito indireto que não desejaríamos (B_2 x B = + 0,005993). Este efeito indireto que nos desvia do alvo é o viés ou tendenciosidade, e a variável AREA responsável por isto é a variável mascarada.

2.3 A ANÁLISE DE CAMINHOS

Conforme Wonnacott (1990), a análise de caminhos é uma generalização do diagrama da figura 1, quando existem mais de duas variáveis independentes.

Se na amostra do nosso exemplo tivesse sido coletada mais uma variável independente, por exemplo a variável topografia do lote, sendo denominada X3 = TOP, conforme apresentada na tabela 2, a regressão múltipla com as variáveis independentes ordenadas da esquerda para a direita em ordem decrescente de causa e efeito (correlação simples com a variável dependente), seria agora dada por:

$$\hat{PU} = 220{,}376 - 0{,}046543\ DIST + 9{,}367045\ TOP - 0{,}069084\ AREA \qquad \text{(eq.2.19)}$$

p-valor (bi-caudal) = 0,0000025 0,00747611 0,00042666

REGISTRO	DISTÂNCIA (m)	TOP	ÁREA (m^2)	PU (R$/$m^2$)
1	2.200,00	0	300,00	100,00
2	2.000,00	1	340,00	110,00
3	1.800,00	0	270,00	120,00
4	1.500,00	1	360,00	140,00
5	2.300,00	0	400,00	85,00
6	1.900,00	1	500,00	105,00
7	1.300,00	0	600,00	120,00
8	2.200,00	0	300,00	95,00
9	900,00	0	360,00	150,00
10	1.700,00	0	600,00	100,00

(TOP=1 para lotes planos e TOP=0 em caso contrário)

Tabela 2 – amostra incluindo a variável TOP

Para determinar os efeitos indiretos da variável DIST em PU via as variáveis AREA e TOP, cada variável independente deve ser tomada como dependente em um modelo de regressão onde as variáveis independentes são as situadas a sua esquerda no modelo de regressão múltipla com todas as variáveis independentes.

Então teremos que determinar os seguintes modelos:

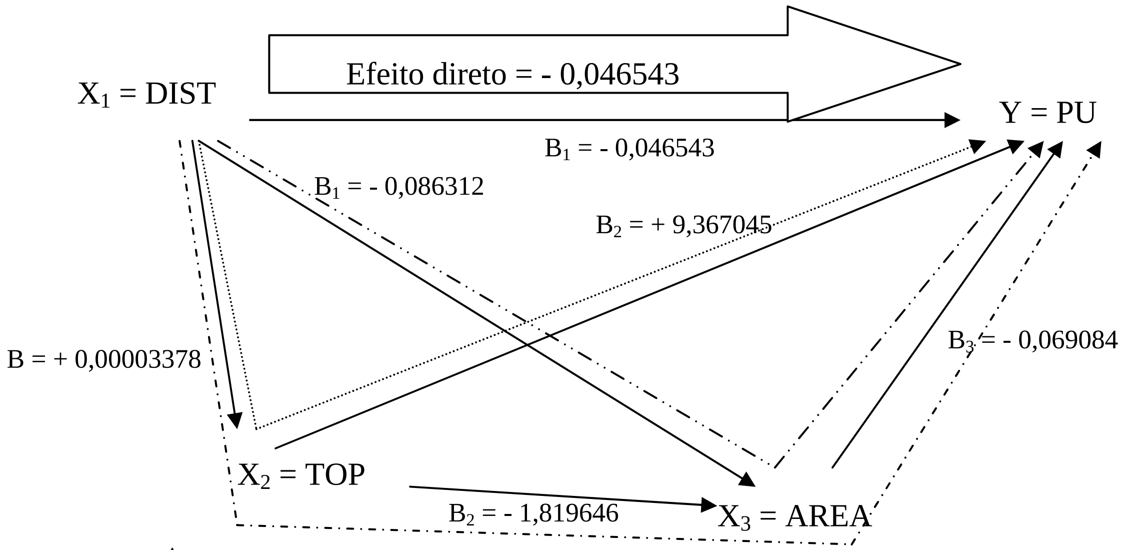
$\widehat{AREA}$ = 557,182 - 0,086312 DIST - 1,819646 TOP (eq.2.20)

p-valor (bi-caudal) = 0,409670 0,98452

$\widehat{TOP}$ = 0,240 + 0,00003378 DIST (eq.2.21)

p-valor (bi-caudal) = 0,932103

Com estas regressões, podemos então registrar os seus parâmetros num diagrama de caminhos semelhante ao da figura 1, mas agora com mais dois caminhos (para chegar em Y partindo de X_1 existem agora 4 caminhos, um direto e três indiretos), conforme a figura 2:



$\widehat{PU}$ = 220,376- 0,046543 DIST + 9,367045 TOP - 0,069084 AREA

$\widehat{AREA}$ = 557,182 - 0,086312 DIST - 1,819646 TOP

$\widehat{TOP}$ = 0,240 + 0,00003378 DIST

$\widehat{PU}$ =184,161 - 0,040259 DIST

Figura 2 – Diagrama de caminhos da regressão com o efeito direto e indireto de DIST

Para calcular o efeito total da variação de DIST em PU, teremos que fazer o somatório do seu efeito direto (mantidas as outras variáveis constantes) e dos seus efeitos indiretos (efeito da variação correspondente das demais variáveis correlacionadas com DIST).

O efeito direto da variação unitária da distância à orla (mantida a topografia e a área) é representado pelo regressor B_{11} da variável distância no modelo da regressão múltipla (B_{11} = - 0,046543).

O efeito indireto da variação unitária da distância à orla via a variação correspondente da variável topografia sozinha é representado pela composição do regressor B_{13} da regressão simples da topografia em função da distância (B_{13} = + 0,00003378) e o regressor B_{21} da variável topografia no modelo da regressão múltipla (B_{21} = + 9,367045), resultando:

B_{13} x B_{21} = (+ 0,00003378) x (+ 9,367045) = + 0,0003164.

O efeito indireto da variação unitária da distância à orla via a variação correspondente da variável topografia e da variável área é representado pela composição do regressor B_{13} da regressão simples da topografia em função da distância (B_{13} = + 0,00003378), com o regressor B_{22} da variável topografia no modelo da regressão múltipla entre a variável dependente AREA em função de DIST e TOP (B_{22} = - 1,819646), com o regressor B_{13} da variável AREA no modelo da regressão múltipla (B_{31} = - 0,069084), resultando:

$B_{13} \times B_{22} \times B_{31}$ = (+ 0,00003378) x (- 1,819646) x (- 0,069084) = + 0,00000425.

O efeito indireto da variação unitária da distância à orla via a variação correspondente da variável AREA sozinha é representado pela composição do regressor B_{12} da variável DIST na regressão múltipla da AREA em função de DIST e TOP (B_{12} = - 0,086312) e o regressor B_{13} da variável área no modelo da regressão múltipla (B_{31} = - 0,069084), resultando:

$B_{12} \times B_{31}$ = (- 0,086312) x (- 0,069084) = + 0,0059628.

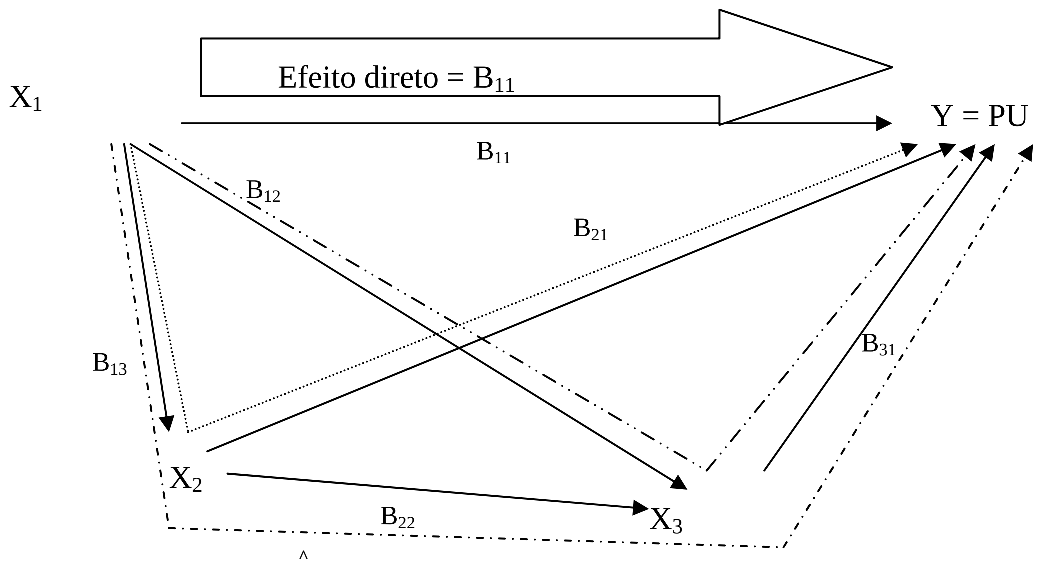
Então, somando os quatro efeitos, resulta:

- 0,046543 + 0,0003164 + 0,00000425 + 0,0059628 = - 0,040259

Este resultado é coincidente com o regressor B_1 da regressão simples do preço unitário em função da distância (B_1 = - 0,040259) obtido na equação 2.14, ou seja:

$$B_{11} + B_{13} \times B_{21} + B_{13} \times B_{22} \times B_{31} + B_{12} \times B_{31} = B_1 \qquad \text{(eq. 2.22)}$$

Resumindo, o efeito total de uma variável independente X_1 na variável dependente Y é definido como a variação que ocorre em Y para uma variação unitária de X_1, levando em conta todas as mudanças nas variáveis intervenientes entre X_1 e Y. Este efeito total pode ser calculado pela soma dois efeitos combinados de todos os caminhos de X_1 para Y, resumido no diagrama de caminhos apresentado na figura 3.



$$\hat{Y} = \mathbf{B_{01} + B_{11} X_1 + B_{21} X_2 + B_{31} X_3}$$

$$\hat{X}_3 = \mathbf{B_{02} + B_{12} X_1 + B_{22} X_2}$$

$$\hat{X}_2 = \mathbf{B_{03} + B_{13} X_1}$$

$$\hat{Y} = \mathbf{B_0 + B_1 X_1}$$

Figura 3 – Diagrama de caminhos da regressão com o efeito direto e indireto de X_1

Para determinar os efeitos indiretos da variável TOP em PU via as variáveis DIST e AREA, cada variável independente deve ser tomada como dependente em um modelo de regressão onde as variáveis independentes são as situadas a sua esquerda no modelo de regressão múltipla com todas as variáveis independentes:

$$\widehat{PU} = 220{,}376 + 9{,}367045\ TOP - 0{,}046543\ DIST - 0{,}069084\ AREA \qquad \text{(eq.2.23)}$$

$$\mathbf{\hat{Y} = B_{01} + B_{11}\ X_1 + B_{21}\ X_2 + B_{31}\ X_3}$$

p-valor (bi-caudal) = 0,00747611 0,0000025 0,00042666

Então teremos que determinar os seguintes modelos:

$$\widehat{AREA} = 557{,}182 - 1{,}819646\ TOP - 0{,}086312\ DIST \qquad \text{(eq.2.24)}$$

$$\mathbf{\hat{X}_3 = B_{02} + B_{12}\ X_1 + B_{22}\ X_2}$$

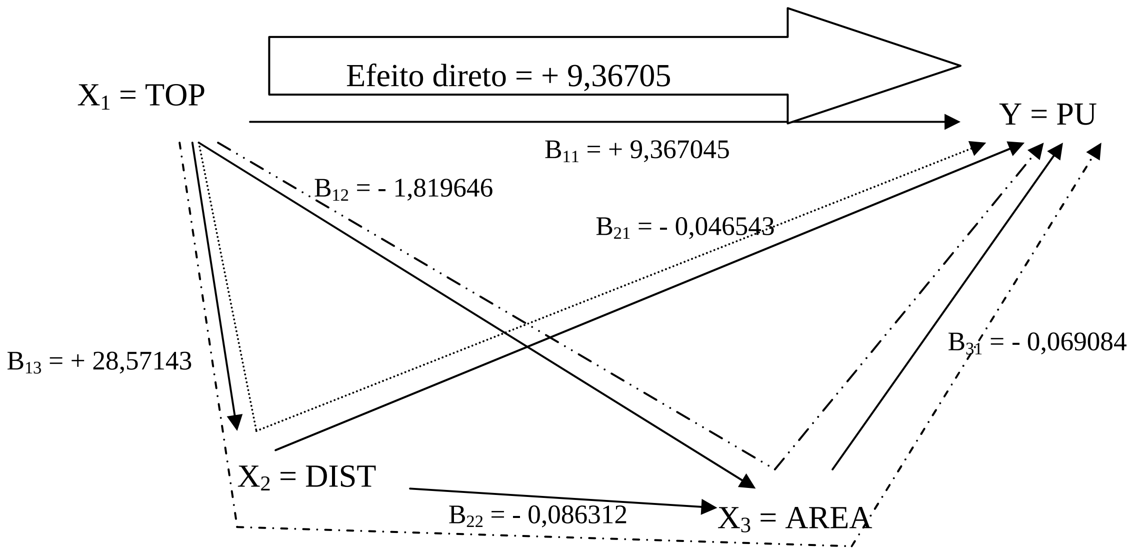
p-valor (bi-caudal) = 0,98452 0,409670

$$\widehat{DIST} = 1771{,}4286 + 28{,}57143\ TOP \qquad \text{(eq.2.25)}$$

$$\mathbf{\hat{X}_2 = B_{03} + B_{13}\ X_1}$$

p-valor (bi-caudal) = 0,93210

Com estas regressões, podemos então registrar os seus parâmetros num diagrama de caminhos semelhante ao da figura 3, conforme a figura 4:



$$\mathbf{\widehat{PU} = 220{,}376 + 9{,}367045\ TOP - 0{,}046543\ DIST - 0{,}069084\ AREA}$$

$$\mathbf{\widehat{AREA} = 557{,}182 - 1{,}819646\ TOP - 0{,}086312\ DIST}$$

$$\mathbf{\widehat{DIST} = 1771{,}4286 + 28{,}57143\ TOP}$$

$$\mathbf{\widehat{PU} = 110 + 8{,}3333\ TOP}$$

Figura 4 – Diagrama de caminhos da regressão com o efeito direto e indireto de TOP

O efeito direto e os três efeitos indiretos de TOP em PU se somam, resultando, pela equação 2.22, o efeito total, que é coincidente com o regressor B_1 da regressão simples do preço unitário em função da topografia (B_1 = 8,3333):

$$B_{11} + B_{13} \times B_{21} + B_{13} \times B_{22} \times B_{31} + B_{12} \times B_{31} = B_1$$

9,36705 + 28,57143 x (- 0,046543) + 28,57143 x (- 0,086312) x (- 0,069084) + (- 1,819646) x (- 0,069084) =

$$9{,}36705 - 1{,}3298 + 0{,}170365 + 0{,}1256708 = 8{,}3333 \qquad \text{(eq.2.26)}$$

Para determinar os efeitos indiretos da variável AREA em PU via as variáveis DIST e TOP, cada variável independente deve ser tomada como dependente em um modelo de regressão onde as variáveis independentes são as situadas a sua esquerda no modelo de regressão múltipla com todas as variáveis independentes:

$$\hat{PU} = 220{,}376 - 0{,}069084\ AREA - 0{,}046543\ DIST + 9{,}367045\ TOP \qquad \text{(eq.2.27)}$$

$$\hat{\mathbf{Y}} = \mathbf{B_{01}} + \mathbf{B_{11}}\ \mathbf{X_1} + \mathbf{B_{21}}\ \mathbf{X_2} + \mathbf{B_{31}}\ \mathbf{X_3}$$

p-valor (bi-caudal) = 0,00042666 0,0000025 0,00747611

Então teremos que determinar os seguintes modelos:

$$\hat{TOP} = = 0{,}257524 - 0{,}0000317178\ AREA + 0{,}00003104419\ DIST \qquad \text{(eq.2.28)}$$

$$\hat{\mathbf{X}}_3 = \mathbf{B_{02}} + \mathbf{B_{12}}\ \mathbf{X_1} + \mathbf{B_{22}}\ \mathbf{X_2}$$

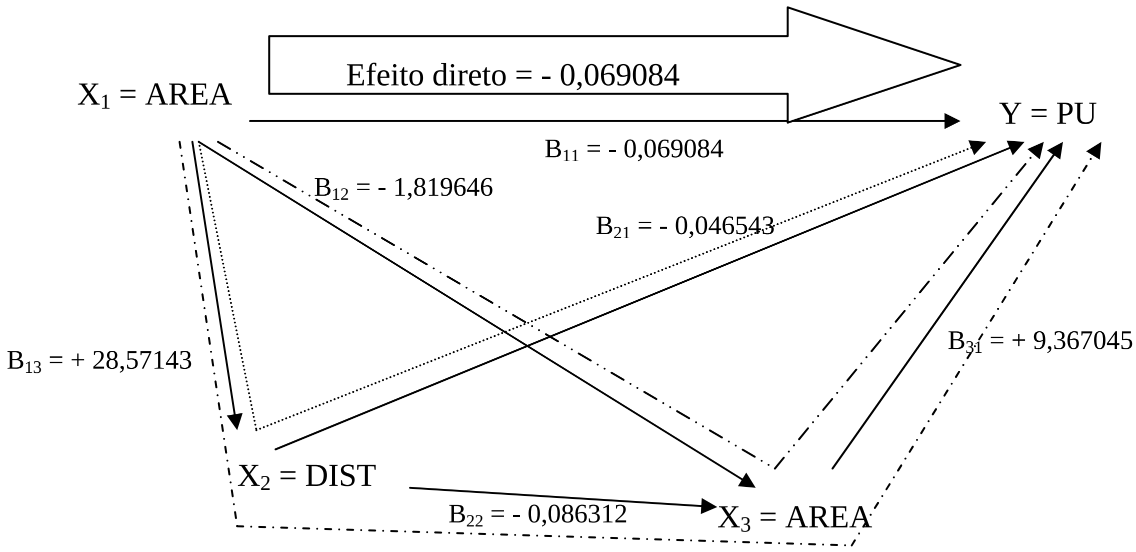
p-valor (bi-caudal) = 0,984524 0,944825

$$\hat{DIST} = 2242{,}6914 - 1{,}1481177\ AREA \qquad \text{(eq.2.29)}$$

$$\hat{\mathbf{X}}_2 = \mathbf{B_{03}} + \mathbf{B_{13}}\ \mathbf{X_1}$$

p-valor (bi-caudal) = 0,37548

Com estas regressões, podemos então registrar os seus parâmetros num diagrama de caminhos semelhante ao da figura 3, conforme a figura 5:



$\widehat{PU}$ = 220,376 - 0,069084 AREA - 0,046543 DIST + 9,367045 TOP

$\widehat{TOP}$ = 0,257524 - 0,0000317178 AREA + 0,00003104419 DIST

$\widehat{DIST}$ = 2242,6914 - 1,1481177 AREA

$\widehat{PU}$ = 119,06032 -0,016278 AREA

Figura 5 – Diagrama de caminhos da regressão com o efeito direto e indireto de AREA

O efeito direto e os três efeitos indiretos de AREA em PU se somam, resultando, pela equação 2.22, o efeito total, que é coincidente com o regressor B_1 da regressão simples do preço unitário em função da área (B_1 = -0,0162787):

$B_{11} + B_{13} \times B_{21} + B_{13} \times B_{22} \times B_{31} + B_{12} \times B_{31} = B_1$

- 0,069084 + (- 1,1481177 **)** x (- 0,046543) + (- 1,1481177 **)** x (0,00003104419) x (**+** 9,367045) + (- 0,0000317178) x 9,36705 =

- 0,069084 + 0,0534368 - 0,00033386 - 0,00029709 = - 0,016278 (eq.2.30)

3. O TRATAMENTO DE HOMOGENEIZAÇÃO POR FATORES

A técnica de homogeneização por meio de fatores consiste em proceder modificações nos preços de cada elemento da amostra, decorrentes da alteração dos diversos atributos, de modo que, ao final do tratamento, os preços homogeneizados se refiram a um imóvel de características padronizadas (paradigma), que poderão até mesmo ser coincidentes com a do imóvel avaliando.

Pode-se formular este tratamento através da seguinte expressão:

$$P_{hom}(i) = \frac{P_{ini}(i)}{F_1(i).F_2(i).F_3(i).............F_k(i)} \quad , i=1,n \qquad \text{(eq. 3.1)}$$

sendo:

P_{hom} (i) = preço homogeneizado do imóvel correspondente ao registro "i" da amostra;
P_{ini} (i) = preço inicial do imóvel correspondente ao registro "i" da amostra;
F_1. (i) = fator de homogeneização relativo ao atributo "1"do registro "i", que espelha a diferença entre o preço do imóvel para o de outro que tenha o atributo "1" igual ao padrão de comparação;
F_2. (i) = fator de homogeneização relativo ao atributo "2" do registro "i", que espelha a diferença entre o preço do imóvel para o de outro que tenha o atributo "2" igual ao padrão de comparação;
F_3. (i) = fator de homogeneização relativo ao atributo "3"do registro "i", que espelha a diferença entre o preço do imóvel para o de outro que tenha o atributo "3" igual ao padrão de comparação;
F_k (i) = fator de homogeneização relativo ao atributo "k" do registro "i", que espelha a diferença entre o preço do imóvel para o de outro que tenha o atributo "k" igual ao padrão de comparação;
n = quantidade de registros na amostra.

Um indicativo do valor do imóvel avaliando ($\hat{P}(a)$) pode ser obtido a partir da média dos preços homogeneizados ($\overline{P}_{hom}$), através da operação inversa, ou seja:

$$\hat{P}(a) = \overline{P}_{hom}.\ F_1(a).F_2(a).F_3(a)........F_k(a) \qquad \text{(eq. 3.2)}$$

sendo:

$$\overline{P}_{hom} = \frac{1}{n}\sum_{i=1}^{n} P_{hom}(i) \qquad \text{(eq. 3.3)}$$

$F_1(a)$, $F_2(a)$, $F_3(a)$, ..., $F_k(a)$ os fatores para o imóvel avaliando

Caso o imóvel avaliando seja o próprio padrão de comparação, $F_1(a)=F_2(a)=F_3(a)=.....= F_k(a) = 1$, e então:

$$\hat{P}(a) = \overline{P}_{hom}.$$

Um dos atributos que podem apresentar diferença entre os diversos registros da amostra é a área do imóvel (A(i)). Caso se entenda que a influência da mesma no preço seja de forma diretamente proporcional, pode-se trabalhar com preços unitários, ou seja:

$$PU_{hom}(i) = \frac{PU_{ini}(i)}{F_1(i).F_2(i).F_3(i).............F_k(i)}, \qquad i=1,n \qquad \text{(eq.3.4)}$$

sendo:

$PU_{hom}(i) = \dfrac{P_{hom}(i)}{A(i)}$ = preço unitário homogeneizado do imóvel correspondente ao registro "i" da amostra;

$PU_{ini}(i) = \dfrac{P_{ini}(i)}{A(i)}$ = preço unitário inicial do imóvel correspondente ao registro "i" da amostra;

Caso se tenha trabalhado com preços unitários, então:

$$\hat{P}(a) = \hat{PU}(a) \, . \, A(a)$$

sendo:

$$\hat{PU}(a) = \overline{PU}_{hom} . \, .F_1(a).F_2(a).F_3(a)........F_k(a) \qquad \text{(eq.3.5)}$$

com $\overline{PU}_{hom} = \dfrac{1}{n}\sum_{i=1}^{n} PU_{hom}(i)$ (eq.3.6)

Caso o paradigma para a homogeneização não tenha sido o imóvel avaliando, se pode, após a definição dos fatores de homogeneização, efetuar uma transposição de paradigma impondo que o produto dos fatores de homogeneização do imóvel avaliando seja igual à unidade e recalculando o produto dos fatores de homogeneização de cada registro da amostra, ou seja:

$$F_{hr}(i) = \frac{F_{ha}(i)}{F_{ha}(a)} \qquad \text{(eq.3.7)}$$

sendo

$$F_{ha}(i) = F_1(i).F_2(i).F_3(i).....F_k(i) \qquad \text{(eq.3.8)}$$

$$F_{ha}(a) = F_1(a).F_2(a).F_3(a)........F_k(a) \qquad \text{(eq.3.9)}$$

A homogeneização é então procedida com o fator de homogeneização relativo F_{hr}, sendo agora o paradigma o imóvel avaliando, ou seja:

$$P_{hom}(i) = \frac{P_{ini}(i)}{F_{hr}(a)} \,, \qquad i=1,n \qquad \text{(eq.3.10)}$$

Daí decorre que a média destes preços homogeneizados será o indicativo do valor do imóvel avaliando $\hat{P}(a)$, ou seja, a equação 3.2 resulta $\hat{P}(a) = \overline{P}_{hom}$ ou, caso se tenha trabalhado com preços unitários, a equação 3.5 resulta $\hat{PU}(a) = \overline{PU}_{hom}$.

3.1. O COEFICIENTE DE HOMOGENEIDADE DOS MODELOS DE HOMOGENEIZAÇÃO POR FATORES.

O mesmo raciocínio que é feito para determinar o indicativo do valor do imóvel avaliando pode ser feito para cada um dos elementos da amostra, bastando considerar que, a cada cálculo, o imóvel avaliando tenha como atributos os de cada registro da amostra.

Com isto se teria para cada registro o seu preço inicial P_{ini} (i) e o seu indicativo de valor determinando pela técnica de homogeneização por fatores $\hat{P}(i)$, com a aplicação da equação 3.2 ou 3.5, conforme os valores sejam totais ou unitários.

A diferença entre P_{ini} (i) e $\hat{P}(i)$ podemos denominar resíduo, não explicado pelo modelo de homogeneização por fatores, ou seja:

$$R(i) = P_{ini}(i) - \hat{P}(i) \qquad \text{(eq. 3.11)}$$

A diferença entre P_{ini} (i) e a média amostral $\overline{P_{ini}}$ se pode denominar variação total inicial (T), não explicada antes da utilização da homogeneização por fatores, ou seja:

$$T(i) = P_{ini}(i) - \overline{P_{ini}} \qquad \text{(eq. 3.12)}$$

A diferença entre $\hat{P}(i)$ e $\overline{P_{ini}}$ se pode denominar variação explicada (E) ao utilizar a homogeneização por fatores, ou seja:

$$E(i) = \hat{P}(i) - \overline{P_{ini}} \qquad \text{(eq.3.13)}$$

Estes três parâmetros podem ser relacionados através da expressão:

$$T(i) = E(i) + R(i) \qquad \text{(eq.3.14)}$$

Em Lima (2001), foi proposto se denominar coeficiente de homogeneidade do modelo (CH) à relação:

$$CH = \frac{\sum_{i=1}^{n} T(i)^2 - \sum_{i=1}^{n} R(i)^2}{\sum_{i=1}^{n} T(i)^2} \qquad \text{(eq.3.15)}$$

Este coeficiente de homogeneidade do modelo assume valor máximo igual a 1, podendo assumir valores negativos.

O valor CH=1 corresponde a que todos os R(i) sejam nulos , ou seja P_{ini} (i) = $\hat{P}(i)$ e o tratamento de homogeneização por fatores levou a estimar os valores de $\hat{P}(i)$ idênticos aos P_{ini} (i) , ou seja, o tratamento de homogeneização por fatores teve o poder de reduzir toda a variação dos preços em relação a média amostral, não resultando nenhum resíduo não explicado.

O valor CH=0 pode corresponder a que todos os T(i) = R(i) e, consequentemente, os E(i) sejam nulos, ou seja todos os $\hat{P}(i) = \overline{P_{ini}}$ e o tratamento de homogeneização por fatores não alterou em nada a alternativa inicial, antes de qualquer tratamento, que seria estimar os $\hat{P}(i)$ pela média dos P_{ini} amostrais ($\overline{P_{ini}}$), ou seja, o tratamento de homogeneização por fatores não teve nenhum poder de diminuição da variação dos preços em relação a média amostral.

O valor CH<0 corresponde a que ao invés de reduzir a variação dos preços em relação à média, o modelo de tratamento de homogeneização por fatores aumentou esta variação, heterogeneizando ao invés de homogeneizar, resultando resíduos não explicados superiores às variações iniciais.

3.2. A ANÁLISE COMPARATIVA ENTRE DOIS TRATAMENTOS DE HOMOGENEIZAÇÃO POR FATORES.

Para uma mesma amostra, a comparação entre os tratamentos pode ser feita pela comparação entre os coeficientes de homogeneidade dos modelos, pois nos dois tratamentos:

- os somatórios dos $T(i)^2$ são idênticos;
- maior coeficiente de homogeneidade implica na menor soma dos quadrados dos resíduos não explicados, ou seja:

$$CH_{máximo} \Rightarrow \left(\sum_{i=1}^{n} R(i)^2\right)_{mínimo} \text{ e o modelo é mais bem ajustado.}$$

Esta comparação dos tratamentos pelo coeficiente de homogeneidade dos modelos pode ser estendida mesmo para o caso de utilização de amostras diferentes, uma vez que neste caso, embora os somatórios dos $T(i)^2$ não sejam idênticos, o maior coeficiente de homogeneidade do modelo implica que a soma dos quadrados dos resíduos não explicados (R(i)) tenha sido percentualmente mais reduzida em relação ao somatório do quadrado dos desvios em relação à média amostral (T(i)).

4. OS FATORES DE HOMOGENEIZAÇÃO FUNDAMENTADOS

Conforme Lima (1995)[6], os fatores de homogeneização fundamentados são oriundos de estudos de mercado através da estatística, utilizando modelos de regressão linear múltipla.

Os fatores são derivados dos regressores obtidos na equação mais ajustada à pesquisa de mercado, do tipo:

$$\ln(\hat{PU}) = B_0 + B_1 X_1 + B_2 X_2 + B_3 X_3 + \dots\dots\dots\dots + B_k X_k \qquad \text{(eq. 4.1)}$$

sendo:

PU = preço unitário,

[6] Ver Lima, Gilson Pereira de Andrade. Homogenheização Fundamentada - Uma Utopia? VIII Congresso Brasileiro de Engenharia de Avaliações e Perícias, Florianópolis, SC, 1995.

B_0 , B_1 , B_2 , B_3 , ... B_k são os regressores obtidos no modelo de regressão,
X_1 , X_2 , X_3 ,................. X_k são os atributos dos imóveis utilizados na homogeneização (variáveis independentes).

A imposição de um modelo cuja variável dependente está na forma transformada logarítmica decorre dos fatores de homogeneização se apresentarem como um acréscimo ou decréscimo percentual nos preços.

Independentemente desta nossa intenção de utilizar o modelo para inferir fatores de homogeneização, em muitos casos onde foi estudado o tipo de equação mais adequado ao comportamento do mercado imobiliário, a transformada logarítmica do preço unitário se mostrou a mais ajustada, como podemos observar em Wolferson (1980), Dantas (1987), Zeni (1990), Chaves Neto (1990), Franchi (1992), Barbosa Filho (1992), Martins (1993) e Gonzáles (1995).

Partindo do modelo de regressão, retroagindo a transformação logarítmica, resulta:

$$\hat{PU} = e^{\left(B_0 + B_1X_1 + B_2X_2 + B_3X_3 + ... + B_kX_k\right)}$$

$$\hat{PU} = e^{B_0} . e^{B_1X_1} . e^{B_2X_2} . e^{B_3X_3} ....e^{B_kX_k} \qquad \text{(eq. 4.2)}$$

E, finalmente :

$$\hat{PU} = P_0 . F_1 . F_2 . F_3 .... F_k \qquad \text{(eq. 4.3)}$$

sendo

$P_0 = e^{B_0}$ = preço unitário do padrão de comparação (paradigma),
com $F_1 = F_2 = F_3 =$ .... $= F_k = 1$

F_1 , F_2 , F_3 , ..., F_k são os fatores de homogeneização derivados do modelo de regressão, e portanto dados por:

$F_1 = e^{B_1X_1}$

$F_2 = e^{B_2X_2}$

$F_3 = e^{B_3X_3}$

....

....

$F_i = e^{B_iX_i}$ (eq. 4.4)

....

....

$F_k = e^{B_kX_k}$

Caso a variável independente (X_j) seja discreta, assumindo portanto valores em quantidade finita, os coeficientes de homogeneização gerados são também em quantidade finita - aplicável à variáveis quantitativas do tipo: quantidade de vagas de garagem, quantidade de dormitórios, etc.-, sendo um mínimo de dois para o caso de variáveis dicotômicas, que por exemplo podem assumir os valores “0” ou “1” - aplicável à variáveis qualitativas tipo: oferta/venda, frente/fundos, etc.).

Caso a variável independente (X_j) seja contínua, assumindo valores diversos, os coeficientes de homogeneização gerados não serão em quantidade finita, sendo função do valor da variável - aplicável a variáveis quantitativas do tipo: idade, índice do local, área, etc.

As variáveis X_1 , X_2 , X_3 ,................. X_k podem estar na forma direta ou transformada (exponencial, inversa, potência ou logarítmica), a que melhor se ajustar ao fenômeno.

A partir dos PU dos “n” dados amostrais e dos “k” fatores de homogeneização determinados, os “n” indicativos de preço do imóvel paradigma (com $F_1 = F_2 = F_3 = \ldots\ldots\ldots\ldots = F_k = 1$), podem ser obtidos pela equação 3.4 e a sua média (eq. 3.6) é o indicativo pontual de valor.

5. A ANÁLISE DE CAMINHOS E OS FATORES DE HOMOGENEIZAÇÃO FUNDAMENTADOS

Se a variável Y do modelo de regressão múltipla é uma transformada logarítmica do preço (P), teremos, pela equação 2.3:

$$\Delta(\ln P) = B_1\,\Delta X_1 + \ldots\ldots + B_i\,\Delta X_i + \ldots\ldots + B_k\,\Delta X_k \qquad \text{(eq. 5.1)}$$

$$(\ln P_{alterado} - \ln P_{original}) = B_1\,\Delta X_1 + \ldots\ldots + B_i\,\Delta X_i + \ldots\ldots + B_k\,\Delta X_k$$

$$\frac{P_{alterado}}{P_{original}} = e^{\left(B_1.\Delta X_1 + B_2.\Delta X_2 + B_3.\Delta X_3 + \ldots\ldots\ldots\ldots\ldots\ldots + B_k.\Delta X_k\right)}$$

$$\frac{P_{alterado}}{P_{original}} = e^{\left(B_1.\Delta X_1\right)} \ldots\ldots\ldots\ldots\, e^{\left(B_i.\Delta X_i\right)} \ldots\ldots\ldots e^{\left(B_k.\Delta X_k\right)}$$

$$\frac{P_{alterado}}{P_{original}} = e^{\left(B_1.X_1 alterado - B_1.X_1 original\right)} \ldots.$$

$$e^{\left(B_i.X_i alterado - B_i.X_i original\right)} \ldots. e^{\left(B_k.X_k alterado - B_k.X_k original\right)}$$

$$\frac{P_{alterado}}{P_{original}} = \left[\frac{e^{\left(B_1.X_1 alterado\right)}}{e^{\left(B_1.X_1 original\right)}}\right] \ldots \left[\frac{e^{\left(B_i.X_i alterado\right)}}{e^{\left(B_i.X_i original\right)}}\right] \ldots. \left[\frac{e^{\left(B_k.X_k alterado\right)}}{e^{\left(B_k.X_k original\right)}}\right] \qquad \text{(eq. 5.2)}$$

Ou seja, a variação percentual no preço é função do produto da variação percentual de cada um dos diversos fatores de homogeneização fundamentados, um para cada uma das variáveis independentes do modelo de regressão.

Se o $P_{alterado}$ é o preço de um imóvel que tem características diferentes do avaliando, sendo este considerado paradigma, e portanto com fator de homogeneização unitário, o indicativo do preço do imóvel avaliando será o $\hat{P}_{original}$ calculado pela equação obtida a partir da equação 5.2, tomando cada fator $e^{(Bi\ Xi\ original)} = 1$, ou seja:

$$\hat{P}_{original} = \left[\frac{P_{alterado}}{e^{\left(B_1.X_1 alterado\right)} \ldots . e^{\left(B_i.X_i alterado\right)} \ldots ; e^{\left(B_k.X_k alterado\right)}} \right] \qquad \text{(eq. 5.3)}$$

Se na homogeneização por fatores forem utilizadas todas as variáveis do modelo de regressão que originou os fatores fundamentados, o modelo de homogeneização por fatores é apropriado, pois, como vimos na análise de caminhos da regressão múltipla apresentada no item 2.3, os parâmetros dos regressores determinados representam a variação da variável dependente Y para uma variação unitária do respectivo regressor, mantidos os demais constantes. Entretanto, se os fatores utilizados se referirem a uma menor quantidade de variáveis independentes que no modelo de regressão que originou os fatores, para que o modelo de homogeneização por fatores seja adequado, os fatores de homogeneização devem ser recalculados para o efeito total, incorporando o efeito indireto das variáveis não consideradas. Mesmo assim, utilizando os fatores recalculados, a omissão de variáveis importantes poderá levar a modelos fracos.

5.1 EXEMPLO

Desenvolvendo-se um exemplo numérico a partir da mesma amostra apresentada na Tabela 2, adotando-se a mesma técnica preconizada por Lima (1995) apresentada no item 4, onde para determinar os fatores de homogeneização fundamentados F_{dist}, F_{top} e F_{area}, tem-se que determinar um modelo de regressão linear do tipo:

$$\ln(\hat{PU}) = B_0 + B_1\ Dist + B_2\ Top + B_3\ Area \qquad \text{(eq. 5.4)}$$

que resulta:

$$\ln(\hat{PU}) = 5{,}6194235 - 0{,}0003979294\ Dist + 0{,}08965253\ Top - 0{,}0005686053\ Area \qquad \text{(eq. 5.5)}$$

p-valor (bi-caudal)= 0,000011 0,013889 0,001980

que é equivalente a :

$$\hat{PU} = 275{,}73\ e^{-0{,}0003979294\ Dist}\ e^{+0{,}08965253\ Top}\ e^{-0{,}0005686053\ Area}$$

ou

$\hat{PU} = 275,73 \times F_{dist} \times F_{top} \times F_{area}$

E portanto:

$F_{dist} = e^{-0,0003979294\ Dist}$ (eq. 5.6)

$F_{top} = e^{+0,08965253\ Top}$ (eq. 5.7)

e

$F_{area} = e^{-0,0005686053\ Area}$ (eq. 5.8)

Caso o interesse fosse determinar somente o fator Fdist, o modelo seria:

$\ln(\hat{PU}) = B_0 + B_1\ Dist$ (eq. 5.9)

resultando:

$\ln(\hat{PU}) = 5,3243596 - 0,0003457879\ Dist$ (eq. 5.10)

p-valor (bi-caudal)= 0,0009043

equivalente a :

$\hat{PU} = 205,28\ e^{-0,00034579Dist} = 205,28\ F_{dist}$

E portanto:

$F_{dist} = e^{-0,00034579Dist}$ (eq. 5.11)

Caso o interesse fosse determinar os fatores F_{top}, o modelo seria:

$\ln(\hat{PU}) = B_0 + B_1\ Top$ (eq. 5.12)

resultando:

$\ln(\hat{PU}) = 4,684641 + 0,08072000\ Top$ (eq.5.13)

p-valor (bi-caudal)= 0,53720

equivalente a :

$\hat{PU} = 108,27\ e^{+0,08072000\ Top}$ ou $\hat{PU} = 108,27 \times F_{Top}$

E portanto:

$F_{Top} = e^{+0,08072000\ Top}$ (eq. 5.15)

Caso o interesse fosse determinar os fatores F_{area}, o modelo seria:

$$\ln(\hat{PU}) = B_0 + B_1 \, Area \qquad \text{(eq. 5.16)}$$

resultando:

$$\ln(\hat{PU}) = 4{,}75632 - 0{,}000117775 \, Area \qquad \text{(eq.5.17)}$$

p-valor (bi-caudal)= 0,82229

equivalente a :

$$\hat{PU} = 116{,}317 \; e^{-0{,}000117775 \; Area} \quad \text{ou} \quad \hat{PU} = 116{,}317 \; x \; F_{area}$$

E portanto:

$$F_{area} = e^{-0{,}000117775 \; Area} \qquad \text{(eq. 5.18)}$$

5.1.1 - MODELO 1 – TRATAMENTO DE HOMOGENEIZAÇÃO POR FATORES FUNDAMENTADOS Fdist Ftop e Farea

Suponha que um profissional tenha que avaliar um lote e dispusesse dos fatores de homogeneização anteriormente calculados e divulgados por uma entidade técnica regional reconhecida e, por acaso, dispusesse também da mesma amostra do quadro 2.

Ele poderia optar por um modelo que considerasse tanto a influência da distância à orla, quanto a topografia e o porte. Este profissional poderia adequadamente optar pelo tratamento de homogeneização pelos fatores F_{dist} fundamentado ($F_{dist} = e^{-0{,}0003979294 \; Dist}$), F_{top} fundamentado ($F_{top} = e^{+0{,}08965253 \; Top}$) e F_{area} fundamentado ($F_{area} = e^{-0{,}0005686053 \; Area}$) determinados pelo modelo de regressão múltipla com as variáveis independentes DIST, TOP e AREA.

Estes fatores seriam calculados para cada registro da amostra, resultando os valores apresentados na tabela a seguir:

REGISTRO	**Distância**	F_{dist}	**Topografia**	F_{top}	**Área**	F_{area}
1	2200	0,4167	0	1,0000	300,00	0,8432
2	2000	0,4512	1	1,0938	340,00	0,8242
3	1800	0,4886	0	1,0000	270,00	0,8577
4	1500	0,5505	1	1,0938	360,00	0,8149
5	2300	0,4004	0	1,0000	400,00	0,7966
6	1900	0,4695	1	1,0938	500,00	0,7525
7	1300	0,5961	0	1,0000	600,00	0,7109
8	2200	0,4167	0	1,0000	300,00	0,8432
9	900	0,6990	0	1,0000	360,00	0,8149
10	1700	0,5084	0	1,0000	600,00	0,7109

Para verificar se o modelo de homogeneização resultante levaria a um grau de precisão adequado, poderia ser determinado o coeficiente de homogeneidade. Para tanto, para cada registro da amostra se faria o cálculo do seu indicativo de valor unitário $\hat{PU}(i)$, e a partir dos valores de PU_{ini} (i) e $\hat{PU}(i)$ poderia determinar pela equação 3.11, para cada registro "i", o resíduo R (i) não explicado pelo modelo de homogeneização por fatores.

A partir da média dos preços unitários iniciais $\overline{PU_{ini}}$ = 112,50, se pode determinar pela equação 3.12, para cada registro "i", a variação total inicial $T_{(i)}$ não explicada antes da utilização da homogeneização por fatores.

A partir destas variações $T_{(i)}$ e $R_{(i)}$ de cada registro, se pode, pela equação 3.15, determinar finalmente o coeficiente de homogeneidade (CH) do modelo de homogeneização por fatores, ou seja:

REGISTRO	**PUini**	$\hat{PU}$	**R**	**R²**	**T**	**T²**
1	100,00	96,91	3,09	9,52	(12,50)	156,25
2	110,00	112,20	(2,20)	4,86	(2,50)	6,25
3	120,00	115,59	4,41	19,44	7,50	56,25
4	140,00	135,36	4,64	21,56	27,50	756,25
5	85,00	87,99	(2,99)	8,91	(27,50)	756,25
6	105,00	106,61	(1,61)	2,58	(7,50)	56,25
7	120,00	116,91	3,09	9,57	7,50	56,25
8	95,00	96,91	(1,91)	3,66	(17,50)	306,25
9	150,00	157,12	(7,12)	50,71	37,50	1.406,25
10	100,00	99,70	0,30	0,09	(12,50)	156,25
Média=>	**112,50**		**Somatório=>**	**130,90**	**Somatório=>**	**3.712,50**

$$CH = \frac{\sum_{i=1}^{n} T(i)^2 - \sum_{i=1}^{n} R(i)^2}{\sum_{i=1}^{n} T(i)^2} = +\ 0{,}9647$$

Este coeficiente de homogeneidade do modelo resultou positivo e razoavelmente próximo de 1, significando que houve uma homogeneização efetiva, pois 96,47% da variabilidade dos preços em relação à média amostral foi reduzida ao aplicar o modelo de homogeneização por fatores, ou seja, ainda restando uma variabilidade residual de 3,53% não explicada pelos fatores de homogeneização F_{dist}, F_{top} e F_{area} utilizados. Isto pode ser comprovado ao observar que a soma dos quadrados dos resíduos finais (R) resultou bem inferior à soma dos quadrados das incertezas iniciais (T) em relação à média amostral.

O gráfico de valores estimados pelo modelo *versus* preços observados pode ser observado na figura 6:

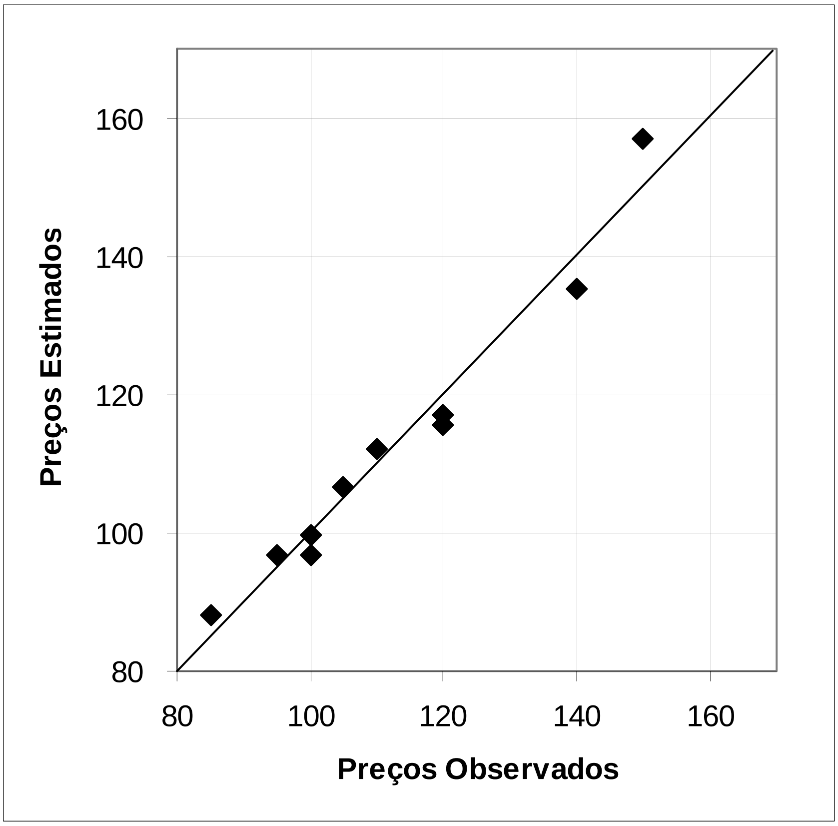


Figura 6 – Gráfico de Preços Estimados do Modelo 1 *versus* Preços Observados

Só para constar, cabe aqui registrar que, aplicando o método de Chauvenet para a eliminação de dados supostamente discrepantes, resultou que, em todas as dez homogeneizações feitas, o método não indica a eliminação de nenhum registro.

5.1.2 - MODELO 2 – TRATAMENTO DE HOMOGENEIZAÇÃO POR FATOR FUNDAMENTADO F_{dist} DETERMINADO NA REGRESSÃO SIMPLES

Supondo que o profissional tenha inicialmente optado por considerar somente a influência da distância à orla. Este profissional poderia adequadamente optar pelo tratamento de homogeneização pelo fator F_{dist} fundamentado ($F_{dist} = e^{0,00034579Dist}$) determinado pelo modelo de regressão simples com a variável independente DIST (eq.5.10), ou poderia calculá-lo a partir dos fatores obtidos na regressão múltipla (eq.5.5) utilizando a análise de caminhos descrita no item 2.3. Este fator seria calculado para cada registro da amostra, resultando os valores apresentados na tabela a seguir:

REGISTRO	DISTÂNCIA	F_{dist}
1	2200	0,4673
2	2000	0,5008
3	1800	0,5366
4	1500	0,5953
5	2300	0,4514
6	1900	0,5184
7	1300	0,6379
8	2200	0,4673
9	900	0,7326
10	1700	0,5555

Da mesma forma como foi feito no modelo 1, para cada registro se refaz o cálculo do seu indicativo de valor unitário, e a partir dos valores de PU_{ini} (i) e $\hat{PU}(i)$ se pode determinar pela equação 3.11, para cada registro "i", o resíduo R (i) não explicado pelo modelo de homogeneização por fatores.

A partir da média dos preços unitários iniciais $\overline{PU_{ini}}$ = 112,50, se pode determinar pela equação 3.12, para cada registro "i", a variação total inicial $T_{(i)}$ não explicada antes da utilização da homogeneização por fatores.

A partir destas variações $T_{(i)}$ e $R_{(i)}$ de cada registro, se pode, pela equação 3.15, determinar finalmente o coeficiente de homogeneidade (CH) do modelo de homogeneização por fatores, ou seja:

REGISTRO	PUini	$\hat{PU}$	R	R^2	T	T^2
1	100,00	96,24	3,76	14,13	(12,50)	156,25
2	110,00	103,13	6,87	47,16	(2,50)	6,25
3	120,00	110,52	9,48	89,91	7,50	56,25
4	140,00	122,60	17,40	302,82	27,50	756,25
5	85,00	92,97	(7,97)	63,53	(27,50)	756,25
6	105,00	106,76	(1,76)	3,10	(7,50)	56,25
7	120,00	131,38	(11,38)	129,43	7,50	56,25
8	95,00	96,24	(1,24)	1,54	(17,50)	306,25
9	150,00	150,87	(0,87)	0,75	37,50	1.406,25
10	100,00	114,41	(14,41)	207,54	(12,50)	156,25
Média=>	**112,50**		**Somatório=>**	**859,91**	**Somatório=>**	**3.712,50**

$$CH = \frac{\sum_{i=1}^{n} T(i)^2 - \sum_{i=1}^{n} R(i)^2}{\sum_{i=1}^{n} T(i)^2} = +\ 0{,}7684$$

Este coeficiente de homogeneidade do modelo resultou positivo e não tão próximo de 1 quanto o do modelo 1, significando que houve uma homogeneização menos efetiva, como era de se esperar pois neste modelo não foram considerados os fatores das variáveis topografia e área, que também são importantes. Agora só 76,84% da variabilidade dos preços em relação à média amostral foi reduzida ao aplicar o modelo de homogeneização por fatores, ou seja, ainda restando uma variabilidade residual de 23,16% não explicada pelo fator de homogeneização F_{dist} utilizado. Isto pode ser comprovado ao observar que a soma dos quadrados dos resíduos finais (R) resultou bem inferior à soma dos quadrados das incertezas iniciais (T) em relação à média amostral.

O gráfico de preços observados *versus* valores estimados pelo modelo pode ser observado na figura 7:

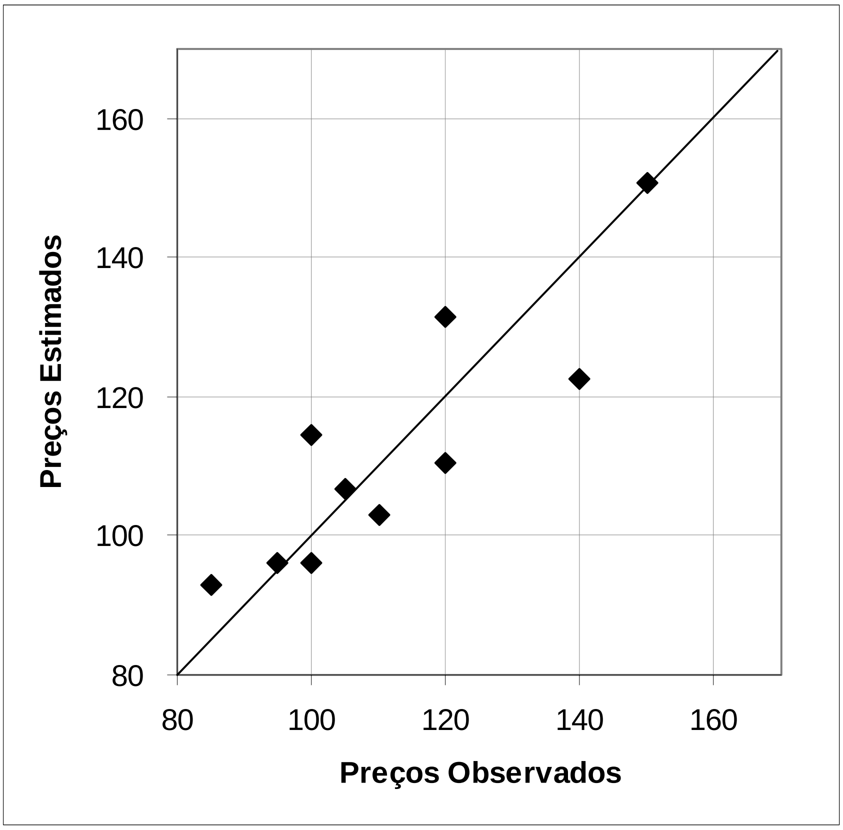


Figura 7 – Gráfico de Preços Estimados do Modelo 2 *versus* Preços Observados

Só para constar cabe aqui registrar que, aplicando o método de Chauvenet para a eliminação de dados supostamente discrepantes, resulta que, em todas as dez homogeneizações feitas, o método não indica a eliminação de nenhum registro.

5.1.3 - MODELO 3 – TRATAMENTO DE HOMOGENEIZAÇÃO POR FATOR FUNDAMENTADO Fdist DETERMINADO NA REGRESSÃO MÚLTIPLA

Supondo que o profissional tenha erroneamente optado por considerar somente a influência da distância à orla, desprezando a influência da topografia e da área, mas utilizando o fator de homogeneização determinado no modelo de regressão múltipla, onde entraram as variáveis distância à orla, topografia e área.. Este profissional utilizaria no tratamento de homogeneização o fator Fdist fundamentado $F_{dist} = e^{-\ 0,0003979294\ Dist}$ (ao invés de Fdist = $e^{\ 0,00034579Dist}$ do modelo 2), sendo este fator calculado para cada registro da amostra, resultando os valores apresentados na tabela a seguir, onde também registramos, para comparação, os fatores do modelo 2:

REGISTRO	DISTÂNCIA	Modelo 3 Fdist = $e^{-0{,}0003979294\,Dist}$	Modelo 2 Fdist = $e^{0{,}00034579 Dist}$
1	2200	0,4167	0,4673
2	2000	0,4512	0,5008
3	1800	0,4886	0,5366
4	1500	0,5505	0,5953
5	2300	0,4004	0,4514
6	1900	0,4695	0,5184
7	1300	0,5961	0,6379
8	2200	0,4167	0,4673
9	900	0,6990	0,7326
10	1700	0,5084	0,5555

Da mesma forma como foi feito no modelo 2, para cada registro se refaz o cálculo do seu indicativo de valor unitário, e a partir dos valores de PU_{ini} (i) e $\hat{PU}(i)$ se pode determinar pela equação 3.11, para cada registro "i", o resíduo $R\,(i)$ não explicado pelo modelo de homogeneização por fatores.

A partir da média dos preços unitários iniciais $\overline{PU_{ini}}$ = 112,50, se pode determinar pela equação 3.12, para cada registro "i", a variação total inicial $T_{(i)}$ não explicada antes da utilização da homogeneização por fatores.

A partir destas variações $T_{(i)}$ e $R_{(i)}$ de cada registro, se pode, pela equação 3.15, determinar finalmente o coeficiente de homogeneidade (CH) do modelo de homogeneização por fatores, ou seja:

REGISTRO	PUini	$\hat{PU}$	R	R^2	T	T^2
1	100,00	94,18	5,82	33,90	(12,50)	156,25
2	110,00	101,98	8,02	64,33	(2,50)	6,25
3	120,00	110,43	9,57	91,64	7,50	56,25
4	140,00	124,43	15,57	242,45	27,50	756,25
5	85,00	90,50	(5,50)	30,29	(27,50)	756,25
6	105,00	106,12	(1,12)	1,25	(7,50)	56,25
7	120,00	134,74	(14,74)	217,17	7,50	56,25
8	95,00	94,18	0,82	0,68	(17,50)	306,25
9	150,00	157,98	(7,98)	63,74	37,50	1.406,25
10	100,00	114,91	(14,91)	222,31	(12,50)	156,25
Média=>	**112,50**		**Somatório=>**	**967,76**	**Somatório=>**	**3.712,50**

$$CH = \frac{\sum_{i=1}^{n} T(i)^2 - \sum_{i=1}^{n} R(i)^2}{\sum_{i=1}^{n} T(i)^2} = +\,0{,}7393$$

Este coeficiente de homogeneidade do modelo resultou positivo e relativamente próximo de 1 mas inferior ao do modelo 2, significando que houve uma homogeneização menos efetiva como era de se esperar, pois neste modelo os fatores adotados não compensaram o viés resultante das variáveis não consideradas (topografia e área). Neste caso 73,93% da variabilidade dos preços em relação à média amostral foi reduzida ao aplicar o modelo de homogeneização por fatores, ou seja, ainda restando uma variabilidade residual de 26,07% não explicada pelos fatores de homogeneização utilizados. Isto pode ser comprovado ao observar que a soma dos quadrados dos resíduos finais (R) resultou bem inferior à soma dos quadrados das incertezas iniciais (T) em relação à média amostral.

O gráfico de preços observados *versus* valores estimados pelo modelo pode ser observado na figura 8:

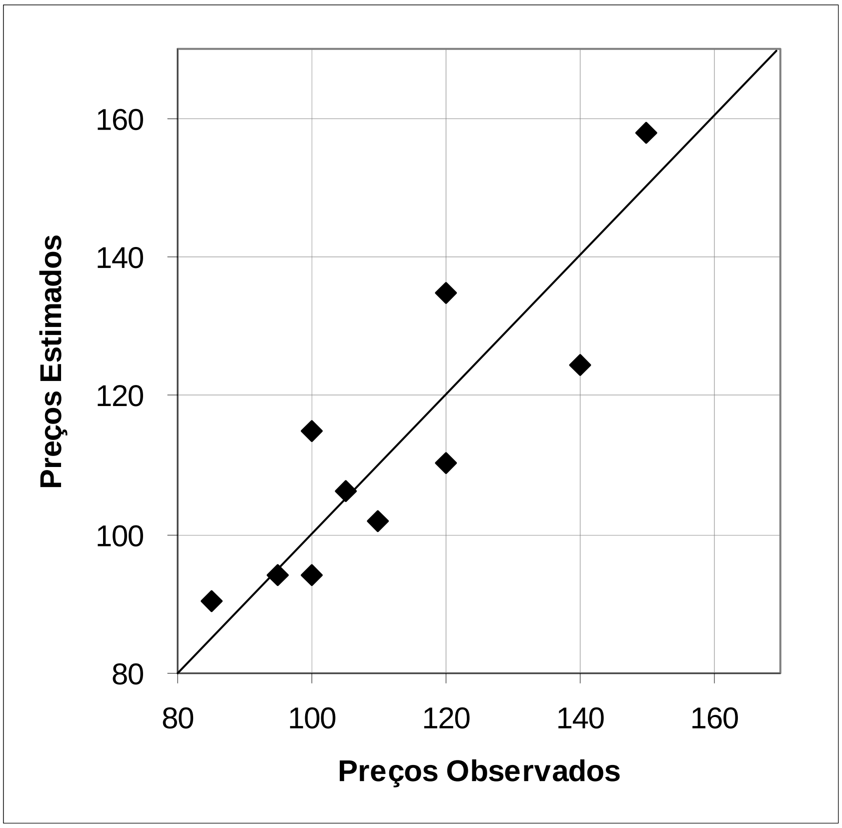


Figura 8 – Gráfico de Preços Estimados do Modelo 3 *versus* Preços Observados

Só para constar, cabe aqui registrar que, aplicando o método de Chauvenet para a eliminação de dados supostamente discrepantes, resulta que, em todas as dez homogeneizações feitas, o método não indica a eliminação de nenhum registro.

5.1.4 - MODELO 4 – TRATAMENTO DE HOMOGENEIZAÇÃO POR FATOR FUNDAMENTADO F_{top} DETERMINADO NA REGRESSÃO SIMPLES

Supondo que o profissional tenha inicialmente optado por considerar somente a influência da topografia. Este profissional poderia adequadamente optar pelo tratamento de homogeneização pelo fator F_{top} fundamentado ($F_{top} = e^{+\ 0,08072000\ Top}$) determinado pelo modelo de regressão simples com a variável independente TOP (eq.5.13), ou poderia calculá-lo a partir dos fatores obtidos na regressão múltipla (eq.5.5) utilizando a análise de caminhos descrita no item 2.3. Este fator seria calculado para cada registro da amostra, resultando os valores apresentados na tabela a seguir:

REGISTRO	Top	F_{top}
1	0	1,0000
2	1	1,0841
3	0	1,0000
4	1	1,0841
5	0	1,0000
6	1	1,0841
7	0	1,0000
8	0	1,0000
9	0	1,0000
10	0	1,0000

Da mesma forma como foi feito no modelo 1, para cada registro se refaz o cálculo do seu indicativo de valor unitário, e a partir dos valores de PU_{ini} (i) e $\hat{PU}(i)$ se pode determinar pela equação 3.11, para cada registro "i", o resíduo $R\,(i)$ não explicado pelo modelo de homogeneização por fatores.

A partir da média dos preços unitários iniciais $\overline{PU_{ini}}$ = 112,50, se pode determinar pela equação 3.12, para cada registro "i", a variação total inicial $T_{(i)}$ não explicada antes da utilização da homogeneização por fatores.

A partir destas variações $T_{(i)}$ e $R_{(i)}$ de cada registro, se pode, pela equação 3.15, determinar finalmente o coeficiente de homogeneidade (CH) do modelo de homogeneização por fatores, ou seja:

REGISTRO	PUini	$\hat{PU}$	R	R²	T	T²
1	100,00	109,75	(9,75)	95,00	(12,50)	156,25
2	110,00	118,97	(8,97)	80,52	(2,50)	6,25
3	120,00	109,75	10,25	105,12	7,50	56,25
4	140,00	118,97	21,03	442,13	27,50	756,25
5	85,00	109,75	(24,75)	612,42	(27,50)	756,25
6	105,00	118,97	(13,97)	195,25	(7,50)	56,25
7	120,00	109,75	10,25	105,12	7,50	56,25
8	95,00	109,75	(14,75)	217,48	(17,50)	306,25
9	150,00	109,75	40,25	1.620,30	37,50	1.406,25
10	100,00	109,75	(9,75)	95,00	(12,50)	156,25
Média=>	**112,50**		**Somatório=>**	**3.568,34**	**Somatório=>**	**3.712,50**

$$CH = \frac{\sum_{i=1}^{n} T(i)^2 - \sum_{i=1}^{n} R(i)^2}{\sum_{i=1}^{n} T(i)^2} = +\,0{,}0388$$

Este coeficiente de homogeneidade do modelo resultou positivo, mas praticamente nulo, significando que houve na realidade uma leve homogeneização. Este resultado, de praticamente nenhuma homogeneização, era de se esperar pois o regressor obtido para a variável TOP no modelo de regressão simples da equação 5.13 foi muito próximo de zero. Neste modelo de homogeneização não foram considerados os fatores da variável distância à orla (o mais importante) e da variável área (de menor importância, mas que também melhoraria o modelo), a não ser de forma indireta, no efeito total da variável topografia obtido na regressão simples.

O gráfico de preços observados *versus* valores estimados pelo modelo pode ser observado na figura 9:

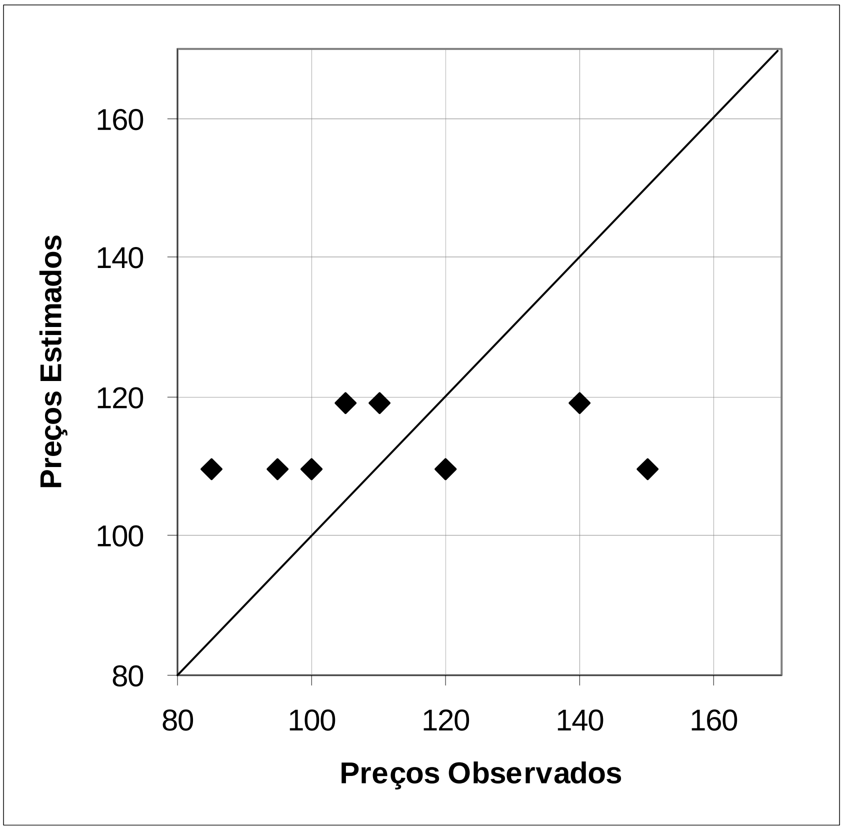


Figura 9 – Gráfico de Preços Estimados do Modelo 4 *versus* Preços Observados

Só para constar cabe aqui registrar que, aplicando o método de Chauvenet para a eliminação de dados supostamente discrepantes, resulta que, em todas as dez homogeneizações feitas, o método só indica a eliminação do registro 9. Eliminando este registro, mais nenhum registro necessita ser eliminado e o coeficiente de homogeneidade do modelo se altera para 0,1740, melhorando o modelo como era de se esperar, mas, entretanto, continuando a ser um modelo muito fraco, pela ausência das demais variáveis já citadas.

5.1.5 - MODELO 5 – TRATAMENTO DE HOMOGENEIZAÇÃO POR FATOR FUNDAMENTADO Ftop DETERMINADO NA REGRESSÃO MÚLTIPLA

Supondo que o profissional tenha erroneamente optado por considerar somente a influência da topografia, desprezando a influência da distância à orla e da área, mas utilizando o fator de homogeneização determinado no modelo de regressão múltipla, onde entraram as variáveis distância à orla, topografia e área. Este profissional utilizaria no tratamento de homogeneização o fator F_{top} fundamentado $F_{top} = e^{+\ 0,08965253\ Top}$ (ao invés de $F_{top} = e^{+0,08072000 Top}$ do modelo 4), sendo este fator calculado para cada registro da amostra, resultando os valores apresentados na tabela a seguir, onde também registramos, para comparação, os fatores do modelo 4:

REGISTRO	Top	Modelo 5 $F_{top} = e^{+\ 0,08965253\ Top}$	Modelo 4 $F_{top} = e^{+0,08072000\ Top}$
1	0	1,0000	1,0000
2	1	1,0938	1,0841
3	0	1,0000	1,0000
4	1	1,0938	1,0841
5	0	1,0000	1,0000
6	1	1,0938	1,0841
7	0	1,0000	1,0000
8	0	1,0000	1,0000
9	0	1,0000	1,0000
10	0	1,0000	1,0000

Da mesma forma como foi feito no modelo 2, para cada registro se refaz o cálculo do seu indicativo de valor unitário, e a partir dos valores de PU_{ini} (i) e $\hat{PU}(i)$ se pode determinar pela equação 3.11, para cada registro "i", o resíduo $R\,(i)$ não explicado pelo modelo de homogeneização por fatores.

A partir da média dos preços unitários iniciais $\overline{PU_{ini}}$ = 112,50, se pode determinar pela equação 3.12, para cada registro "i", a variação total inicial $T_{(i)}$ não explicada antes da utilização da homogeneização por fatores.

A partir destas variações $T_{(i)}$ e $R_{(i)}$ de cada registro, se pode, pela equação 3.15, determinar finalmente o coeficiente de homogeneidade (CH) do modelo de homogeneização por fatores, ou seja:

REGISTRO	PUini	$\hat{PU}$	R	R²	T	T²
1	100,00	109,46	(9,46)	89,41	(12,50)	156,25
2	110,00	119,72	(9,72)	94,52	(2,50)	6,25
3	120,00	109,46	10,54	111,18	7,50	56,25
4	140,00	119,72	20,28	411,19	27,50	756,25
5	85,00	109,46	(24,46)	598,09	(27,50)	756,25
6	105,00	119,72	(14,72)	216,74	(7,50)	56,25
7	120,00	109,46	10,54	111,18	7,50	56,25
8	95,00	109,46	(14,46)	208,97	(17,50)	306,25
9	150,00	109,46	40,54	1.643,83	37,50	1.406,25
10	100,00	109,46	(9,46)	89,41	(12,50)	156,25
Média=>	**112,50**		**Somatório=>**	**3.574,53**	**Somatório=>**	**3.712,50**

$$CH = \frac{\sum_{i=1}^{n} T(i)^2 - \sum_{i=1}^{n} R(i)^2}{\sum_{i=1}^{n} T(i)^2} = +\ 0{,}0371$$

Este coeficiente de homogeneidade do modelo resultou positivo, mas praticamente nulo, significando que houve na realidade uma leve homogeneização. Este resultado, de praticamente nenhuma homogeneização, era de se esperar pois o regressor obtido para a variável TOP no modelo de regressão múltipla da equação 5.5 foi muito próximo de zero. Neste modelo de homogeneização não foram considerados os fatores da variável distância à orla (o mais importante) e da variável área (de menor importância, mas que também melhoraria o modelo), nem mesmo de forma indireta, como no efeito total da variável topografia obtido na regressão simples utilizado no modelo 4, daí que o seu coeficiente de homogeneidade ainda resultou inferior ao do modelo 4.

O gráfico de preços observados *versus* valores estimados pelo modelo pode ser observado na figura 10:

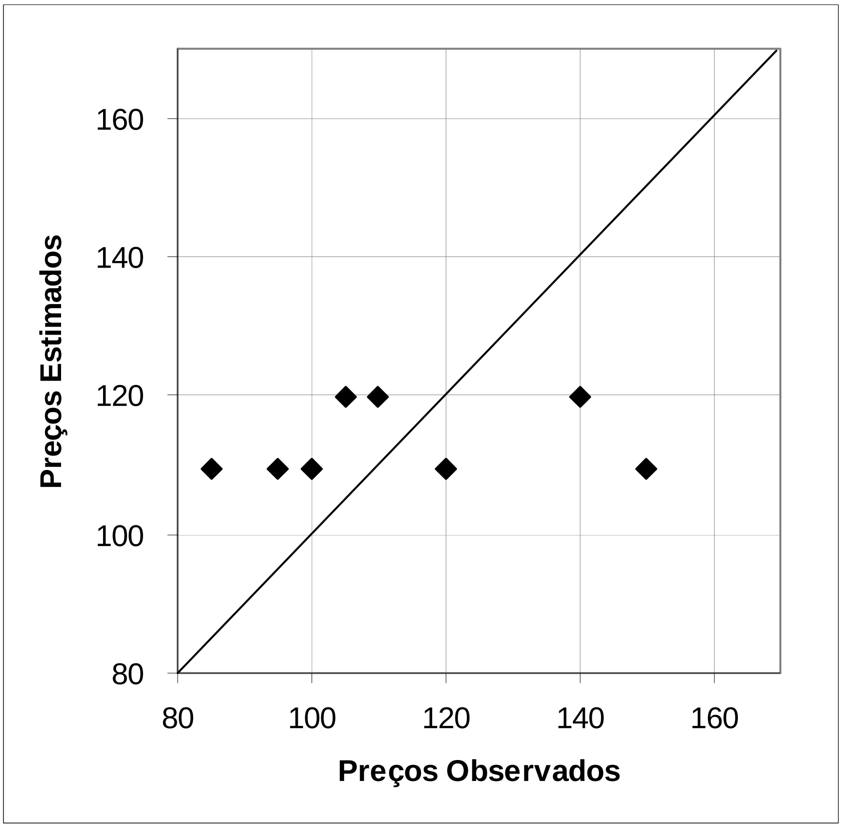


Figura 10 – Gráfico de Preços Estimados do Modelo 5 *versus* Preços Observados

Só para constar cabe aqui registrar que, aplicando o método de Chauvenet para a eliminação de dados supostamente discrepantes, resulta que, em todas as dez homogeneizações feitas, o método só indica a eliminação do registro 9. Eliminando este registro, mais nenhum registro necessita ser eliminado e o coeficiente de homogeneidade do modelo se altera para 0,1845, melhorando o modelo como era de se esperar, mas, entretanto, continuando a ser um modelo muito fraco, pela ausência das demais variáveis já citadas.

5.1.6 - MODELO 6 – TRATAMENTO DE HOMOGENEIZAÇÃO POR FATOR FUNDAMENTADO F_{area} DETERMINADO NA REGRESSÃO SIMPLES

Supondo que o profissional tenha inicialmente optado por considerar somente a influência da área. Este profissional poderia adequadamente optar pelo tratamento de homogeneização pelo fator F_{area} fundamentado ($F_{area} = e^{-\ 0{,}000117775\ Area}$) determinado pelo modelo de regressão simples com a variável independente AREA (eq.5.17), ou poderia calcula-lo a partir dos fatores obtidos na regressão múltipla (eq.5.5) utilizando a análise de caminhos descrita no item 2.3. Este fator seria calculado para cada registro da amostra, resultando os valores apresentados na tabela a seguir:

REGISTRO	Área	F_{area}
1	300,00	0,9653
2	340,00	0,9607
3	270,00	0,9687
4	360,00	0,9585
5	400,00	0,9540
6	500,00	0,9428
7	600,00	0,9318
8	300,00	0,9653
9	360,00	0,9585
10	600,00	0,9318

Da mesma forma como foi feito no modelo 1, para cada registro se refaz o cálculo do seu indicativo de valor unitário, e a partir dos valores de PU_{ini} (i) e $\hat{PU}(i)$ se pode determinar pela equação 3.11, para cada registro "i", o resíduo $R\,(i)$ não explicado pelo modelo de homogeneização por fatores.

A partir da média dos preços unitários iniciais $\overline{PU_{ini}}$ = 112,50, se pode determinar pela equação 3.12, para cada registro "i", a variação total inicial $T_{(i)}$ não explicada antes da utilização da homogeneização por fatores.

A partir destas variações $T_{(i)}$ e $R_{(i)}$ de cada registro, se pode, pela equação 3.15, determinar finalmente o coeficiente de homogeneidade (CH) do modelo de homogeneização por fatores, ou seja:

REGISTRO	PUini	$\hat{PU}$	R	R^2	T	T^2
1	100,00	113,86	(13,86)	192,03	(12,50)	156,25
2	110,00	113,32	(3,32)	11,04	(2,50)	6,25
3	120,00	114,26	5,74	32,94	7,50	56,25
4	140,00	113,06	26,94	725,99	27,50	756,25
5	85,00	112,52	(27,52)	757,59	(27,50)	756,25
6	105,00	111,21	(6,21)	38,53	(7,50)	56,25
7	120,00	109,90	10,10	101,91	7,50	56,25
8	95,00	113,86	(18,86)	355,60	(17,50)	306,25
9	150,00	113,06	36,94	1.364,88	37,50	1.406,25
10	100,00	109,90	(9,90)	98,11	(12,50)	156,25
Média=>	**112,50**		**Somatório=>**	**3.678,62**	**Somatório=>**	**3.712,50**

$$CH = \frac{\sum_{i=1}^{n} T(i)^2 - \sum_{i=1}^{n} R(i)^2}{\sum_{i=1}^{n} T(i)^2} = +\,0{,}0091$$

Este coeficiente de homogeneidade do modelo resultou positivo, mas praticamente nulo, significando que houve na realidade uma leve homogeneização. Este resultado, de praticamente nenhuma homogeneização, era de se esperar pois o regressor obtido para a variável AREA no modelo de regressão simples da equação 5.17 foi muito próximo de zero. Neste modelo de homogeneização não foram considerados os fatores da variável distância à orla (o mais importante) e da variável topografia (de menor importância, mas que também melhoraria o modelo), a não ser de forma indireta, no efeito total da variável área obtido na regressão simples.

O gráfico de preços observados *versus* valores estimados pelo modelo pode ser observado na figura 11:

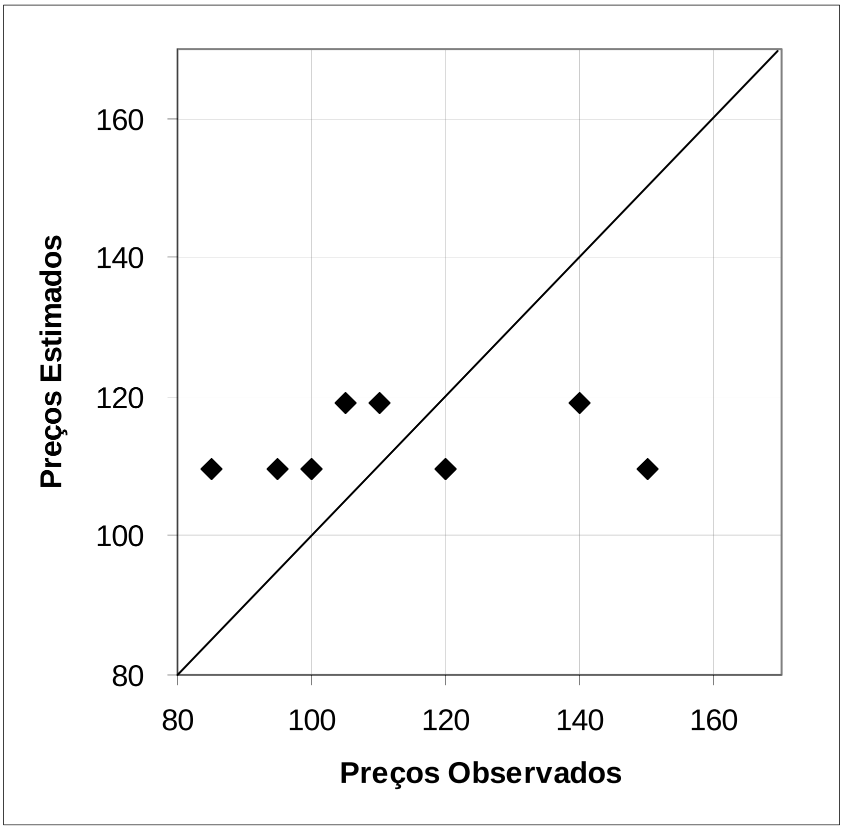


Figura 11 – Gráfico de Preços Estimados do Modelo 6 *versus* Preços Observados

Só para constar cabe aqui registrar que, aplicando o método de Chauvenet para a eliminação de dados supostamente discrepantes, resulta que, em todas as dez homogeneizações feitas, o método não indica a eliminação de nenhum registro.

5.1.7 - MODELO 7 – TRATAMENTO DE HOMOGENEIZAÇÃO POR FATOR FUNDAMENTADO F_{area} DETERMINADO NA REGRESSÃO MÚLTIPLA

Supondo que o profissional tenha erroneamente optado por considerar somente a influência da área, desprezando a influência da distância à orla e da área, mas utilizando o fator de homogeneização determinado no modelo de regressão múltipla, onde entraram as variáveis distância à orla, topografia e área. Este profissional utilizaria no tratamento de homogeneização o fator F_{area} fundamentado $F_{area} = e^{-\ 0,0005686053\ Area}$ (ao invés de $F_{area} = e^{-\ 0,000117775\ Área}$ do modelo 6), sendo este fator calculado para cada registro da amostra, resultando os valores apresentados na tabela a seguir, onde também registramos, para comparação, os fatores do modelo 6:

REGISTRO	Área	Modelo 7 $F_{area} = e^{-0,0005686053\,Area}$	Modelo 6 $F_{area} = e^{-0,000117775}$
1	300,00	0,8432	0,9653
2	340,00	0,8242	0,9607
3	270,00	0,8577	0,9687
4	360,00	0,8149	0,9585
5	400,00	0,7966	0,9540
6	500,00	0,7525	0,9428
7	600,00	0,7109	0,9318
8	300,00	0,8432	0,9653
9	360,00	0,8149	0,9585
10	600,00	0,7109	0,9318

Da mesma forma como foi feito no modelo 2, para cada registro se refaz o cálculo do seu indicativo de valor unitário, e a partir dos valores de PU_{ini} (i) e $\hat{PU}(i)$ se pode determinar pela equação 3.11, para cada registro "i", o resíduo R (i) não explicado pelo modelo de homogeneização por fatores.

A partir da média dos preços unitários iniciais $\overline{PU_{ini}}$ = 112,50, se pode determinar pela equação 3.12, para cada registro "i", a variação total inicial $T_{(i)}$ não explicada antes da utilização da homogeneização por fatores.

A partir destas variações $T_{(i)}$ e $R_{(i)}$ de cada registro, se pode, pela equação 3.15, determinar finalmente o coeficiente de homogeneidade (CH) do modelo de homogeneização por fatores, ou seja:

REGISTRO	PUini	$\hat{PU}$	R	R^2	T	T^2
1	100,00	119,41	(19,41)	376,76	(12,50)	156,25
2	110,00	116,73	(6,73)	45,23	(2,50)	6,25
3	120,00	121,46	(1,46)	2,15	7,50	56,25
4	140,00	115,41	24,59	604,90	27,50	756,25
5	85,00	112,81	(27,81)	773,40	(27,50)	756,25
6	105,00	106,57	(1,57)	2,48	(7,50)	56,25
7	120,00	100,68	19,32	373,12	7,50	56,25
8	95,00	119,41	(24,41)	595,87	(17,50)	306,25
9	150,00	115,41	34,59	1.196,80	37,50	1.406,25
10	100,00	100,68	(0,68)	0,47	(12,50)	156,25
Média=>	**112,50**		**Somatório=>**	**3.971,16**	**Somatório=>**	**3.712,50**

$$CH = \frac{\sum_{i=1}^{n} T(i)^2 - \sum_{i=1}^{n} R(i)^2}{\sum_{i=1}^{n} T(i)^2} = -\ 0{,}0697$$

Este coeficiente de homogeneidade do modelo resultou negativo, mas praticamente nulo, significando que houve na realidade uma leve heterogeneização. Este resultado, de praticamente nenhuma homogeneização, era de se esperar pois o regressor obtido para a variável AREA no modelo de regressão múltipla da equação 5.5 foi muito próximo de zero. Neste modelo de homogeneização não foram considerados os fatores da variável distância à orla (o mais importante) e da variável topografia (de menor importância, mas que também melhoraria o modelo), nem mesmo de forma indireta, como no efeito total da variável área obtido na regressão simples utilizado no modelo 6, daí que o seu coeficiente de homogeneidade ainda resultou inferior ao do modelo 6.

O gráfico de preços observados *versus* valores estimados pelo modelo pode ser observado na figura 12:

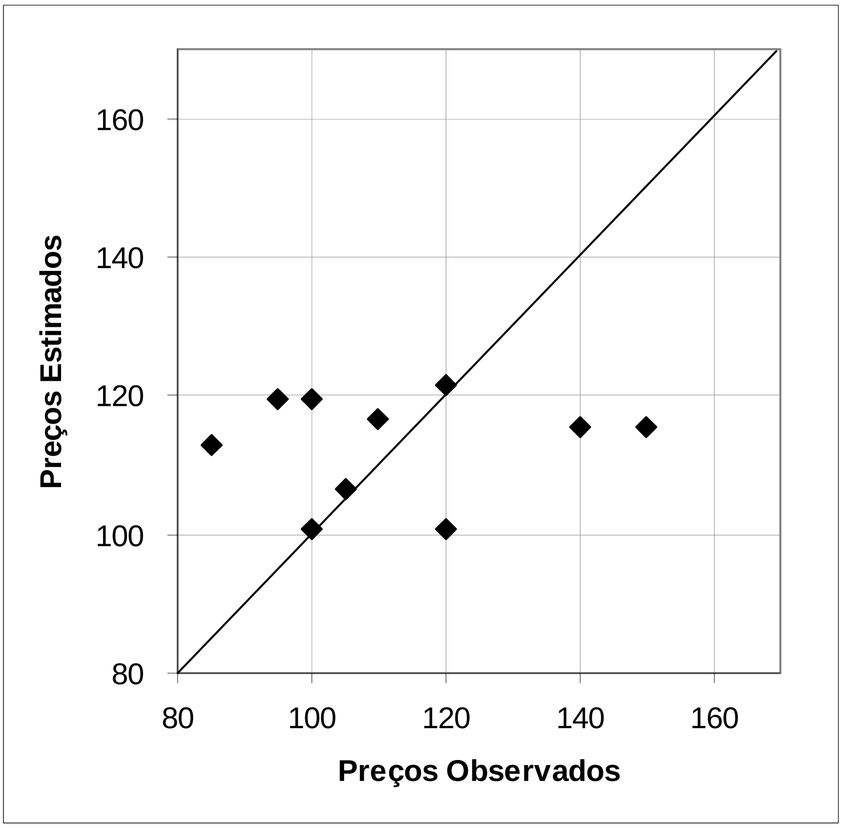


Figura 12 – Gráfico de Preços Estimados do Modelo 7 *versus* Preços Observados

Só para constar cabe aqui registrar que, aplicando o método de Chauvenet para a eliminação de dados supostamente discrepantes, resulta que, em todas as dez homogeneizações feitas, o método não indica a eliminação de nenhum registro.

5.2.4 COMPARAÇÃO DOS COEFICIENTES DE HOMOGENEIDADE ENTRE OS MODELOS DE HOMOGENEIZAÇÃO POR FATORES

Em resumo, se têm os seguintes CH calculados para os sete modelos de homogeneização por fatores:

Modelo	CH
Modelo 1 - Fatores Fundamentados F_{dist} F_{top} e F_{area} da regressão múltipla	+ 0,9647
Modelo 2 - Fatores Fundamentados F_{dist} da regressão simples	+ 0,7684
Modelo 3 – Fator Fundamentado F_{dist} da regressão múltipla	+ 0,7393
Modelo 4 - Fator Fundamentado F_{top} da regressão simples	+ 0,0388
Modelo 5 - Fator Fundamentado F_{top} da regressão múltipla	+ 0,0371
Modelo 6 - Fator Fundamentado F_{area} da regressão simples	+ 0,0091
Modelo 7 - Fator Fundamentado F_{area} da regressão múltipla	- 0,0697

Ou seja, neste exemplo, pelo critério do coeficiente de homogeneidade do modelo(CH), podemos tirar as seguintes conclusões:

- a utilização de fatores fundamentados F_{dist} F_{top} e F_{area} da regressão múltipla resultou no melhor modelo, como era de se esperar;
- o modelo com os fatores fundamentados F_{dist} ou F_{top} ou F_{area}.obtidos da regressão múltipla utilizados isoladamente resultou em modelos inferiores aos das mesmas variáveis F_{dist} ou F_{top} ou F_{area}.obtidos das regressões simples com estas variáveis independentes, porque neste caso estão incorporados os efeitos indiretos das demais variáveis desconsideradas;
- a incorporação dos efeitos indiretos via obtenção dos fatores a partir da regressão simples ou da utilização da análise de caminhos a partir dos fatores da regressão múltipla por si só não compensa plenamente variáveis importantes não consideradas no modelo de homogeneização, como se pode observar dos coeficientes de homogeneidade obtidos nos modelos 4, 5, 6, e 7, onde foi omitida a variável distância à orla.

6 – CONCLUSÕES E RECOMENDAÇÔES

A utilização de modelos de homogeneização por fatores requer cuidados especiais.

Mesmo quando os fatores de homogeneização são fundamentados, obtidos a partir de estudos de mercado via a utilização de modelos de regressão, a omissão de uma variável no modelo de homogeneização pode levar a um modelo inadequado, principalmente se a variável omitida for importante na explicação da variabilidade dos preços.

Na divulgação de coeficientes de homogeneização é importante explicitar que variáveis entraram no modelo e os seus respectivos graus de importância, pela significância dos respectivos regressores.

Seria também interessante que fossem divulgados os regressores dos modelos entre as variáveis independentes de modo a permitir, pela análise de caminhos, a incorporação de efeitos indiretos de variáveis não consideradas no modelo de homogeneização por fatores.

Espera-se que este trabalho auxilie na correta utilização de modelos de homogeneização por fatores, assim como incentive a pesquisa e a correta divulgação de fatores de homogeneização fundamentados a partir da utilização de bancos de dados de registros amostrais de pesquisas de mercado, em muitos casos já disponíveis em diversas instituições.

7. BIBLIOGRAFIA

ABNT (Associação Brasileira de Normas Técnicas). Norma NBR14653-1:2001 - AVALIAÇÃO DE BENS - PARTE 1: PROCEDIMENTOS GERAIS, ABNT, 2001.

Barbosa Filho, Domingos de Saboya. Examinando os modelos de regressões. Caderno Brasileiro de Avaliações e Perícias, Avalien, Porto Alegre, julho, 1992.

Chaves Neto, Raymundo L.V. Comportamento setorizado do mercado de locações — uma análise fundamentada em modelos de regressão. VI Congresso Brasileiro de Engenharia de Avaliações e Perícias, Belo Horizonte, 1990.

Dantas, Rubens e Cordeiro, Gauss Moutinho. A avaliação de imóveis através da metodologia de pesquisa científica. Melhor trabalho apresentado no V Congresso Brasileiro de Engenharia de Avaliações e Perícias - COBREAP - laureado com a medalha Eurico Ribeiro, Recife, PE, 1987. . Caderno Brasileiro de Avaliações e Perícias, Avalien, Porto Alegre, julho, agosto, 1991.

Franchi, Cláudia de Cesare. Avaliação das características que contribuem para a formação do valor de apartamentos na cidade de Porto Alegre. Dissertação de Mestrado apresentada ao corpo docente do Curso de Pós-graduação em Engenharia Civil da Escola de Engenharia da Universidade Federal do Rio Grande do Sul, Caderno Brasileiro de Avaliações e Perícias, Avalien, Porto Alegre, abril, maio e junho, 1992.

González, Marco Aurélio Spunf e Formoso, Carlos Torres. Análise da utilização de dados do imposto de transmissão de imóveis para atualização das plantas de valores. Anais do 1º Congresso Brasileiro de Avaliações para fins tributários, Cachoeira do Sul, 1995.

González, Marco Aurélio Spunf. Determinação de planta de valores com base em dados do ITBI— estudo da viabilidade de aplicação em Porto Alegre. Relatório final que sintetiza a pesquisa, UFRGS, Porto Alegre, 1995.

Lima, Gilson Pereira de Andrade. Homogenheização Fundamentada - Uma Utopia? VIII Congresso Brasileiro de Engenharia de Avaliações e Perícias, Florianópolis, SC, 1995.

Lima, Gilson P.de A., Separando o Joio do Trigo: A Importância da Análise de Resíduos na Homogeneização por Fatores, Anais do 1º Seminário Internacional de Real State da América Latina da LARES – Latin American Real State Society, São Paulo, SP, 1999.

Lima, Gilson P.de A., O coeficiente de Homogeneidade na Homogeneização por Fatores, Anais do 2º Seminário Internacional de Real State da América Latina da LARES – Latin American Real State Society, São Paulo, SP, 2001.

Martins, Fernando Guilherme e Martins, Fábio Guilherme Neuber. Contribuição de Melhoria. Caderno Brasileiro de Avaliações e Perícias, Avalien, Porto Alegre, fevereiro, 1993.

Zeni, André Maciel. Valorização de terrenos na malha urbana — um perfil de formação. VI Congresso Brasileiro de Engenharia de Avaliações e Perícias, Belo Horizonte, 1990.

Wolferson, Marco e Torres, Luiz. Homogeneização de valores em engenharia de avaliações. Recife, ENAPEL, 1980.

Wonnacott, Ronald J. e Wonnacott, Thomas H. Introductory Statistics for Business and Economics. USA, John Wiley & Sons, 4th. Ed. 1990.