

PAINEL: Avaliação de Imóveis - Análise Quantitativa

Do mau uso da inferência estatística aplicada às avaliações imobiliárias - Teoria e Exemplos

Misusing of statistical inference in the real estate appraisal. Theory and examples
Del mal uso de la inferência estadística aplicada a las evaluaciones inmobiliarias-
Teoria y ejemplos

Eng. Paulo Grandiski
Profissional autônomo - São Paulo
e-mail: grandiski@ip2.com.br

RESUMO:

A inferência estatística é atualmente uma das mais usadas ferramentas de trabalho para complexos estudos econométricos de comportamento de vários mercados. As várias metodologias aplicáveis ao mercado imobiliário, exigem a correta montagem de modelos, bem como a admissão de pressupostos, nem sempre aplicáveis, quer pela insuficiência de dados de mercado disponíveis, quer pelos diferentes comportamentos de seus vários sub-segmentos. Por esses motivos, a aplicação da inferência estatística exige dos pesquisadores conhecimento aprofundado tanto dos mercados imobiliários quanto das metodologias empregadas, que lamentavelmente a maior parte dos profissionais não domina. A facilidade do aprendizado dos softwares específicos que surgiram dá aos avaliadores a falsa sensação de que dominam a teoria, transformando-os em simples “*pilotos automáticos de computadores*”. Este trabalho, utilizando exercícios elementares de modelos determinísticos, pretende evidenciar alguns exemplos de mau uso da inferência estatística aplicada às avaliações imobiliárias, tais como:

- falhas na montagem de modelos estatísticos, incorrendo em micronumerosidade ou “excesso de ajustamento”;
- falhas na identificação das principais variáveis que explicam a tendência de formação de preços de mercado, confundindo estas com variáveis secundárias que apenas explicam a variação dos preços de mercado, com base na variação dessas variáveis;
- falhas na identificação das interações entre variáveis, mesmo nos casos em que não há relações de causa-efeito entre elas.

Misusing of statistical inference in the real estate appraisal. Theory and examples

ABSTRACT:

Statistical inference, at the present, is one of the most used tools for complex econometric study work of markets behavior. Many applied methodologies to the real estate market requires the correct models setting up, and the conjecture of hypothesis admission, not every time applied, by datum insufficiency from many sub-segments, or by different behavior of the sub-segments. Because of these reasons, the statistical inference application requires from the researches a deep knowledge of real estate markets and applied methodologies, that lamentably the valuer's major part don't dominate. The apprentice's facility of the specific softwares that had appeared give to the valuers the false sensation that they have the domination of the theory, changing them to simple "*automatic computer's pilot*". This study, with the use of elemental exercises of deterministic models, intends to evidence some examples of misusing of statistical inferences applied to the real estate appraisals, such as:

- setting up error in the statistical model's, incurring in the problem of micronumerosity or "overfitting adjustment";
- fails in the identification of the principal variables that explain the market prices formation tendency, confusing them with the secondary variables that only explain the variation of the market prices, according with the variations of these variables.
- fails in the identification of the variable's interaction, even in the cases where there isn't cause-effect within them.

Del mal uso de la inferencia estadística aplicada a las evaluaciones inmobiliarias – Teoría y ejemplos

RESUMEN:

La inferencia estadística es actualmente una de las herramientas de trabajo mas usadas para complejos estudios econométricos de comportamiento de varios mercados. Las varias metodologías aplicables al mercado inmobiliario, exigen el correcto montaje de modelos, tal como la admisión de presupuestos, no siempre aplicables, sea por la insuficiencia de datos disponibles del mercado, sea por los diferentes comportamientos de sus varios subsegmentos.

Por estos motivos, la aplicación de la inferencia estadística exige de los investigadores un profundo conocimiento, tanto de los mercados inmobiliarios cuanto de las metodologías empleadas, que infelizmente la mayor parte de los profesionales no domina. La facilidad del aprendizaje de los softwares específicos que surgieron, ofrece a los evaluadores la falsa sensación de que dominan la teoría, transformándolos en simples "pilotos automáticos de computadores".

Este trabajo, utilizando ejercicios elementares de modelos determinantes, pretende comprobar algunos ejemplos del mal uso de la inferencia estadística aplicada a las evaluaciones inmobiliarias, tales como:

- fallas en el montaje de los modelos estadísticos, incurriendo en micro numerosidad o "exceso de ajustamiento";
- fallas en la identificación de las principales variables que explican la tendencia de formación de precios del mercado, confundiendo estas con variables secundarias, que apenas explican la variación de los precios del mercado, con base en la variación de estas variables;
- fallas en la identificación de las interacciones entre variables, aún en los casos en que no hay relaciones de causa y efecto entre ellas.

1 - INTRODUÇÃO: OS MERCADOS IMOBILIÁRIOS

Leigos falam em mercado imobiliário como se fosse um único mercado de produtos denominados “imóveis”, bem definidos e homogêneos. Na verdade existem vários mercados imobiliários – sub-segmentos, tais como os mercados de terrenos nus - imóveis rurais, chácaras de lazer, glebas urbanizáveis, terrenos urbanos (para construção de casas isoladas, de conjunto de casas, para incorporações, para indústrias), e os mercados de terrenos com benfeitorias, que podem se diferenciar quanto a padrões arquitetônicos, construtivos, idade, estado de conservação, etc.

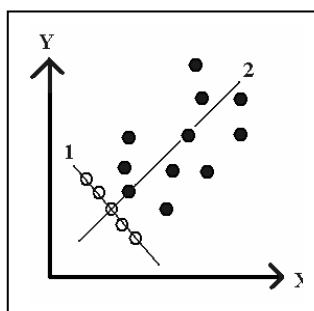
Cada um desses sub-mercados pode apresentar ao longo do tempo tendências de formação de preços substancialmente diferentes entre si, com características atípicas, que TAMBÉM podem mudar de local para local. Por esse motivo, TODAS as diferenças e diferentes comportamentos deveriam constar como variáveis explicativas na modelagem, o que nem sempre é possível no mercado brasileiro, devido à pequena disponibilidade de dados confiáveis de transações contemporâneas e semelhantes ao imóvel em avaliação.

Conseqüentemente, estudos do mercado imobiliário podem ser feitos com base em diferentes características e em várias escalas de detalhamento, com o uso de séries temporais ou através de “cross-sections”, disto resultando margens de erro muito variáveis entre si, para cada tipo de estudo.

Se o estudo do mercado estiver sendo feito com base na teoria cartesiana de tratamento por fatores, não há dúvidas quanto ao fato de as margens de erro serem menores à medida que o foco seja concentrado em cada sub-segmento de mercado, onde é possível encontrar elementos transacionados semelhantes ao elemento avaliando, exigindo apenas pequenos ajustes para compatibilizar as características dos elementos amostrais às respectivas características do elemento avaliando.

Não obstante, a notória dificuldade em encontrar elementos amostrais recentemente transacionados implica rotineiramente na utilização de elementos que, embora com características físicas semelhantes, sejam mais distantes fisicamente do elemento avaliando, e, nestes casos, podem surgir outras variáveis que deveriam ser levadas em conta, sob pena de aumentar os vieses estatísticos.

2 - MICRONEUMEROSIDADE X “EXCESSO DE AJUSTAMENTO”



É recomendável que o uso da inferência estatística para estudo dos mercados imobiliários utilize número de elementos amostrais suficientes para indicar ao modelo o real comportamento de mercado, evitando assim os riscos decorrentes da micronumerosidade, ou seja, utilização de pequeno número de elementos amostrais, nem sempre representativos da população, conforme pode ser visto no desenho ao lado, onde a regressão 1, indicada pelos 5 pontos iniciais, na verdade deveria ser representada pela regressão 2, que fica evidenciada quando se acrescenta maior número de elementos amostrais.

Por outro lado, o **excesso de ajustamento** do modelo de regressão aos elementos amostrais, embora produza coeficientes de determinação (R^2) mais altos, faz com que o modelo fique muito dependente da condição “*ceteri paribus*”. Em outras palavras, se a amostra não for bem representativa da população, ou, pior ainda, se a amostra contiver elementos influenciadores e/ou “*outliers*”, o modelo encontrado conduzirá a resultados muito diferentes com a retirada de alguns

dos pontos influenciantes, que podem ser identificados por apresentarem “**distâncias de Cook**” aumentadas.

Muitos estudos nacionais e internacionais do mercado imobiliário indicam a tendência de uma grande parcela percentual dos preços transacionados de elementos semelhantes de um sub-segmento de mercado, num dado local, e numa certa data, convergirem através de raciocínios lógicos para um valor médio (ou modal) definido pela regra universal:

$$\text{Preço total transacionado} = \text{área do imóvel} \times \text{valor unitário médio}$$

Conseqüentemente, outra parte percentual, não desprezível dos preços transacionados, é representada por variações aleatórias, decorrentes de razões não lógicas, baseada na **emoção** de compradores e vendedores. Esta situação decorre do fato de o mercado imobiliário estar longe de comportamento homogêneo, podendo-se afirmar que não existem sequer dois elementos totalmente iguais entre si: mesmo que existissem máquinas de fabricar casas, elas teriam localizações diferentes, em pavimentos diferentes, insolações diferentes, diferentes estados de conservação, etc.

3 - PREÇOS: TENDÊNCIA DE FORMAÇÃO X TENDÊNCIA DE VARIAÇÃO DE PREÇOS (LOTEAMENTOS TIPO ALPHAVILLE)

Existem no Brasil vários loteamentos padronizados com lotes muito semelhantes, quase todos planos e retangulares, onde todos os lotes têm a mesma testada T, a mesma profundidade P e portanto a mesma área de terreno $S = T \times P$.

Esses loteamentos “fechados”, como o pioneiro Alphaville, na cidade de São Paulo, costumam ter muros divisórios, com entrada única, através de portaria controlada 24h/dia e sistema de segurança motorizada.

Nesses loteamentos, é fácil identificar, num certo instante, um valor unitário médio V_U , que conduz a um preço total médio dos lotes equivalente a

$$V_{T\text{médio}} = S \times V_U \quad (1)$$

Esta fórmula representa a tendência da formação do preço médio desses lotes, em torno do qual os preços dos demais lotes oscilam, para mais ou para menos, por vários motivos, entre os quais, numa visão mais abrangente, podem ser distinguidos três grandes grupos:

1) grupo das variáveis endógenas do mercado de terrenos local, tais como:

- a) se o terreno é em corte ou aterro, com conseqüências diretas no custo da fundação e indiretas no partido arquitetônico (muros de arrimo, número de pavimentos, linda vista, etc.)
- b) a maior ou menor proximidade (distância) do lote à portaria, considerado como fator de maior segurança ostensiva, e portanto, com menor risco de assaltos;
- c) a maior ou menor proximidade (distância) do lote aos muros das divisas, que, teoricamente, constituem lotes com maior risco de assalto.

2) grupo das variáveis exógenas ao mercado de lotes de terreno

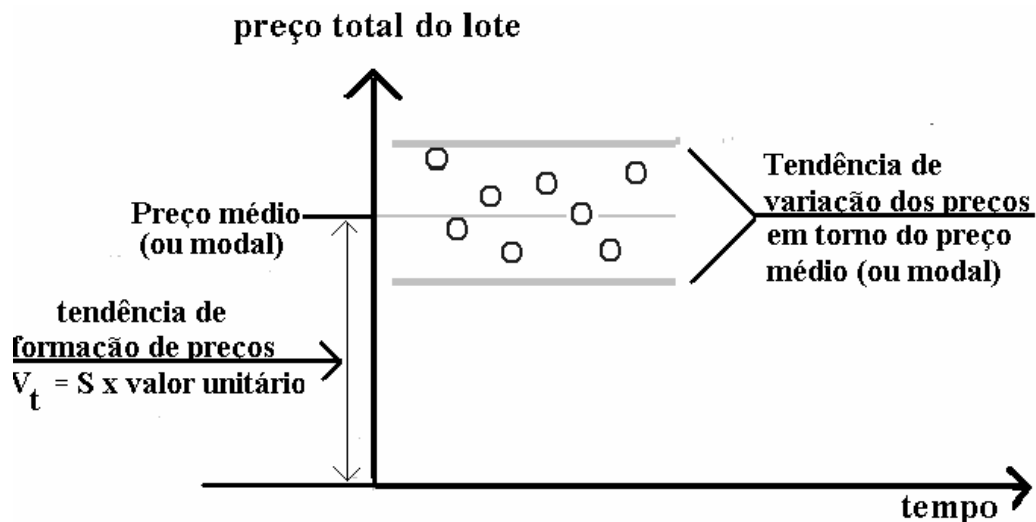
- a) influências (restrições) da legislação local (Planos diretores, lei de zoneamento, Código de Edificações, operações urbanas especiais, etc.

b) influências das oscilações do mercado financeiro, da situação política nacional e internacional, etc.

3) grupo das variações aleatórias (emocionais) altamente subjetivas dos compradores

Estas variações de preço, além das diferentes habilidades de negociação entre compradores e vendedores, incluem desde a disponibilidade de tempo para pesquisas no local, até interesses específicos dos vendedores (por exemplo, urgência em vender) e compradores (interesse em comprar lote vizinho ao que já possui).

Assim sendo, o gráfico de transações recentes desses lotes pode ser representado da seguinte forma:



As equações de regressão representativas do comportamento de mercado podem ser apresentadas na forma de somatórios do tipo

$$Y = a + b_1 \times X_1 + b_2 \times X_2 + b_3 \times X_3 + \dots + b_n \times X_n$$

nas quais Y representa a variável explicada VALOR TOTAL, e X_1 até X_n são as variáveis explicativas, com seus respectivos coeficientes b_1 até b_n

Nos casos de terrenos semelhantes ao loteamento ALPHAVILLE o intercepto “a” costuma representar a verdadeira tendência de formação de preços médios $V_t = S \times \text{valor unitário}$, podendo representar percentuais que variam de 60% a 90% do valor total Y, ou seja, dividindo-se o valor do intercepto “a” pelo valor calculado de Y, se obtém relações a/Y da ordem de 0,6 a 0,9. Em outras palavras, todas as demais n variáveis explicativas, em conjunto, conseguem explicar de 10% a 40% **da variação** dos valores de Y.

CONCLUSÃO 1: Equações de regressão que melhor se aproximam da tendência de formação dos preços dos imóveis devem apresentar baixas relações a/Y , pois no caso de terrenos com valores nulos das variáveis explicativas, o valor de Y deveria também ser nulo. Assim sendo, elas podem representar as verdadeiras tendências de **FORMAÇÃO** de preços, com base na variação das respectivas variáveis explicativas utilizadas.

CONCLUSÃO 2: As equações de regressão que indicam valores de a/Y em grandes percentuais na verdade explicam a **VARIAÇÃO** dos preços em torno do valor médio da região, sendo equivocado confundi-las com as verdadeiras tendências de formação de preço médio. Nestas equações, como a variável área não varia, e os preços variam, os avaliadores mal informados concluem que “a variável área não é importante na formação dos preços”, quando na verdade, ela não é importante

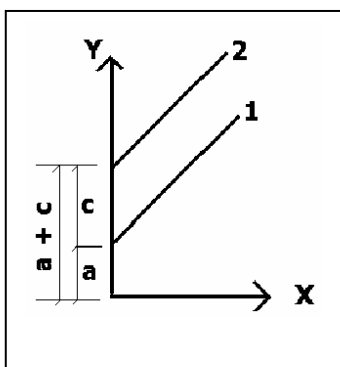
apenas para explicar a variação dos preços, pois ela faz parte do intercepto, que na verdade corresponde ao maior componente da equação dos somatórios de preços.

4 - VARIÁVEIS DICOTÔMICAS (“DUMMIES”) DE INTERCEPTO E DE INCLINAÇÃO

As variáveis dicotômicas costumam ser utilizadas nos modelos de regressão para distinguir duas situações fáticas existentes, tais como “bairro 1” x “bairro 2”, “rua asfaltada” x “rua de terra”, “rua principal” x “rua secundária”, “apartamento com vaga de estacionamento” x “apartamento sem vaga de estacionamento”, e assim por diante.

Rotineiramente essas variáveis dicotômicas costumam utilizar os valores “zero” e “um”, indicando a existência ou não da situação contemplada. Para exemplificar, considere-se o caso simples de uma reta de regressão num plano onde estejam representadas apenas a variável explicada e uma variável explicativa. Neste caso, a variável “dummy” pode indicar variações no intercepto ou variações de inclinação de retas de regressão, como é detalhado a seguir.

4.1 - VARIÁVEIS “DUMMY” DE INTERCEPTO



No esquema ao lado estão representadas as retas de regressão representativas de dois agrupamentos de dados, por exemplo, dados do bairro 1 e dados do bairro 2, nas quais a variável explicativa área tem o mesmo ângulo de inclinação, e conseqüentemente a mesma tangente desse ângulo, que pode equivaler ao mesmo valor unitário do imóvel por metro quadrado, variando apenas a taxa fixa “a” de corretagem, para o bairro 1, e a taxa fixa “a+c” de corretagem para o bairro 2.

A equação de regressão do bairro 1 pode ser indicada por

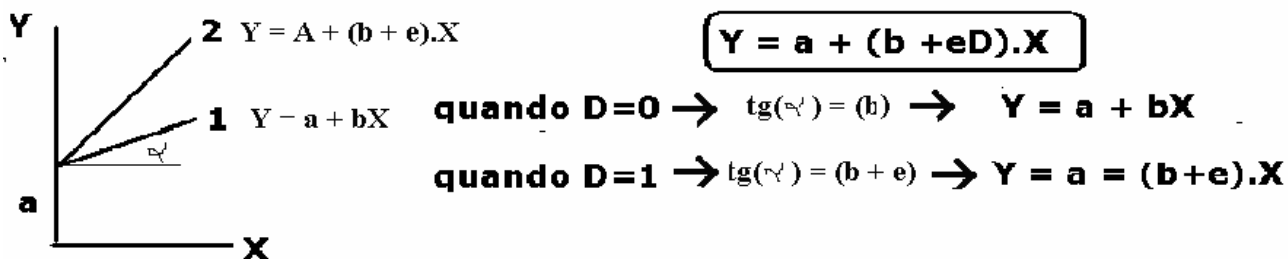
$$\text{Valor total} = Y_1 = a + b.\text{área}$$

Em conseqüência, para representar **na mesma equação** o comportamento dos dois bairros, devemos usar uma variável dummy “D” de intercepto, ficando o modelo de representação dos dois bairros representado por

$$\text{Valor total} = Y = (a + cD) + b.X$$

Quando a variável D assumir o valor zero, indicando que estamos avaliando no bairro 1, o valor correspondente a “c vezes zero” dá zero, e o valor do intercepto fica igual a “a”. Caso contrário, se estivermos avaliando no bairro 2, a variável “dummy” D assume o valor 1, e o valor correspondente a “c vezes um” assume o valor c, que, somado ao valor “a” anterior, forma o intercepto “a+c” do bairro 2.

4.2 VARIÁVEIS “DUMMY” DE INCLINAÇÃO



Neste outro caso os dois agrupamentos de dados dos bairros 1 e 2 apresentam retas de regressão com o mesmo intercepto “a”, mas tangentes dos ângulos de inclinação diferenciadas, e, portanto, a variável “dummy” deverá indicar a variação das inclinações das duas retas de regressão representativas do comportamento desses dois agrupamentos de dados. Se a variável explicativa X continuar a ser a variável “área do terreno”, no bairro 1 o preço unitário médio será “b”, e no bairro 2 poderá ser representado pelo preço médio “(b + e.D)”, podendo os dois bairros ser representados simultaneamente pela equação

$$\text{Valor total} = Y = a + (b + e.D). X$$

Quando estivermos avaliando no bairro 1, a variável dummy D assumirá o valor zero, e portanto a parcela (e.D) terá valor nulo, ficando a tangente do ângulo da reta representado por b, e a tangente do ângulo da reta do bairro 2 representada por (b + e), vez que a variável dummy D assume o valor “1” no bairro 2.

Desmembrando a equação acima com dummy representativa da mudança das tangentes dos ângulos das respectivas retas representativas do comportamento de mercado, resulta

$$\text{Valor total} = Y = a + (b + e.D). X$$

$$\text{Valor total} = Y = a + b.X + e.D.X$$

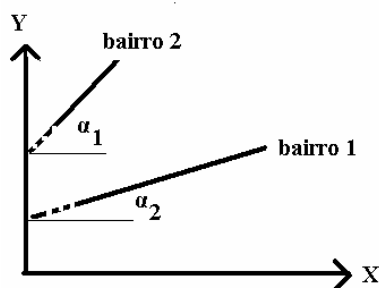
Observe atentamente esta última equação: quando a variável dummy indica as variações das inclinações das retas, surge “automaticamente” na equação unificada uma nova variável “D.X” resultante da multiplicação da variável dummy X com a variável explicativa cuja mudança de inclinação estamos estudando, no caso, a variável explicativa X.

4.3 DIFERENÇAS SIMULTÂNEAS DE INTERCEPTO E DE INCLINAÇÃO ENTRE DOIS AGRUPAMENTOS

Até aqui V. já aprendeu que, misturando elementos amostrais de dois agrupamentos de dados, numa mesma equação de regressão:

- quando os interceptos das retas diferem, devemos incluir uma variável dummy de intercepto;
- quando os ângulos das retas diferem, devemos incluir no modelo uma variável de interação, que corresponde à multiplicação da variável dummy com a variável cujo ângulo se altera.

Quando ocorrerem simultaneamente as diferenças de intercepto e de inclinação entre os dois agrupamentos, a mesma variável “dummy” pode ser empregada simultaneamente para indicar as variações, como demonstrado esquematicamente a seguir:



$$Y = (a + cD) + (b + eD)X$$

$$Y = a + cD + eD + bX + eDX$$

$$Y = a + (c+e)D + bX + eDX$$

$$Y = \underbrace{a + bX}_{\text{equação da reta}} + \underbrace{fD}_{\text{Dummy intercepto}} + \underbrace{eDX}_{\text{Variável de interação}}$$

Esse fato prova que podem estar incidindo em equívoco os avaliadores que tentam obter equações de regressão de hiperplanos representativos de vários bairros de uma cidade (plantas de valores genéricos), que misturam bairros de terrenos super-valorizados com terrenos de favelas, nas quais as interações, inclusive espaciais, deveriam ser investigadas.

Em menor escala o mesmo pode ocorrer quando se tenta distinguir valores de oferta e de transação numa única equação de regressão, usando apenas uma variável dummy de intercepto, conforme fica provado no seguinte exemplo:

Suponha que a equação de regressão de valores ofertados seja
OFERTA: $Y = a + bX$

Admitindo que o fator oferta do mercado seja F, resulta

$$\text{TRANSAÇÃO} = F \cdot \text{OFERTA} = F \cdot (a + bX)$$

$$\text{TRANSAÇÃO} = Fa + Fb \cdot x$$

Observe que entre os dois agrupamentos de dados, oferta e transação, o intercepto mudou de “a” para “F.a”, e que a tangente do ângulo da reta de regressão mudou de “b” para “F.b”. Portanto, quando se quiser representar, numa mesma equação, dados de oferta e dados de transação, não basta introduzir uma dummy de intercepto para o agrupamento de dados de oferta, mas é preciso testar se é significativa a introdução da variável de interação correspondente à multiplicação da variável dummy com a variável OFERTA. Dependendo do fator F, essa interação pode se mostrar significativa, e deve ser levada em conta.

5 – CONCLUSÕES SOBRE AS VARIÁVEIS DE INTERAÇÃO

As interações entre variáveis já eram conhecidas e divulgadas em 1914, quando GOSSET, ainda usando o pseudônimo de “Student”, publicou na seção “MISCELLANEA” da revista *Econometrika* um artigo com o nome “IV – THE ELIMINATION OF SPURIOUS CORRELATION DUE TO POSITION IN TIME OR SPACE”.

No Brasil as interações foram divulgadas, entre os avaliadores, pelo premiado trabalho do eng. Rubens Alves Dantas, que foi apresentado no IX Congresso Brasileiro de Engenharia de

Avaliações realizado em São Paulo, em 1997, sob o título “Avaliação de Imóveis utilizando Modelos Especiais”, e posteriormente detalhados no seu livro “Engenharia de Avaliações – Uma introdução à metodologia científica”

Oito anos depois, em 2005, a maioria dos avaliadores, tanto no Brasil como no Exterior, ainda não estuda a influência dessas interações, deixando de levá-las em conta nos modelos de regressão.

Pior ainda, costumam misturar elementos amostrais pertencentes a mercados de diferentes comportamentos, não apenas em uma variável explicativa, mas em várias.

Nesses casos, conforme detalhado por Damodar N. Gujarati, no Capítulo XV de seu livro “Econometria Básica”, deveriam ser investigadas as interações entre as variáveis “dummies” e TODAS as demais variáveis quantitativas, que podem influenciar os resultados finais, inclusive em razão dos resultantes fenômenos de multicolinearidade e de autocorrelações, inclusive espaciais, que daí podem resultar.

A nova variável multiplicativa que “automaticamente” resulta do fato da utilização de agrupamentos de dados que diferem quanto às inclinações das respectivas retas de regressão, conforme detalhado no item 4.2 acima, costuma ser chamada de “variável de interação” na literatura econométrica.

Na opinião deste autor, essa nomenclatura é imprópria, pois ela tanto pode resultar de reais interações entre variáveis como pode surgir de modelos espúrios, que misturam elementos em nada semelhantes, sem nenhuma “interação” entre elas, ou nenhuma relação de causa-efeito, como infelizmente tem ocorrido entre os avaliadores brasileiros.

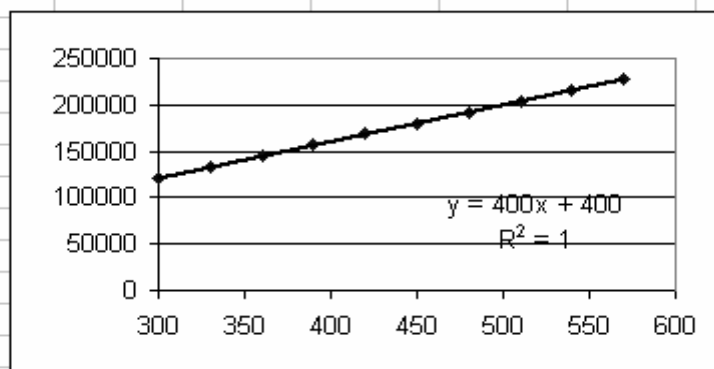
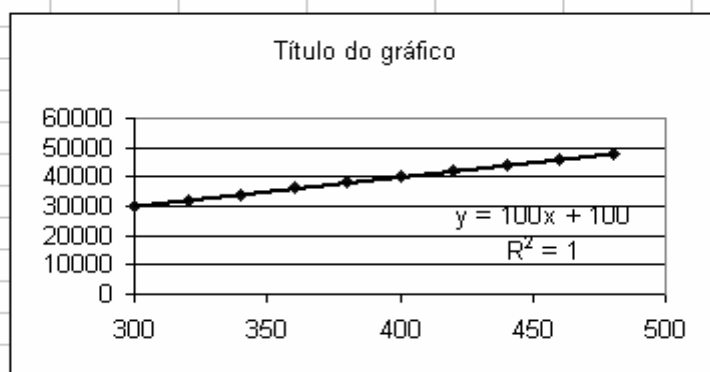
Chamar a atenção para esses fatos foi o principal motivo da apresentação deste estudo, que faz parte de uma série de artigos já publicados por este autor, sob o título “Aperfeiçoando as Avaliações, nos quais alguns aspectos relevantes do mau uso da inferência estatística são comentados.

Para enfatizar os aspectos comentados neste estudo, são apresentados a seguir exemplos muito simples, com dados determinísticos, com os quais o autor demonstra os grandes equívocos que podem resultar do não uso das variáveis de interação, tentando convencer os que utilizam a inferência estatística da necessidade de estudos mais aprofundados da matéria.

6 - EXERCÍCIOS

6.1 - TABELA ORIGINAL DE DADOS DE DOIS BAIROS, NA FORMA DETERMINÍSTICA

ELEMENTO	área	BAIRRO	valor total
1	300	0	30100
2	320	0	32100
3	340	0	34100
4	360	0	36100
5	380	0	38100
6	400	0	40100
7	420	0	42100
8	440	0	44100
9	460	0	46100
10	480	0	48100
11	300	1	120400
12	330	1	132400
13	360	1	144400
14	390	1	156400
15	420	1	168400
16	450	1	180400
17	480	1	192400
18	510	1	204400
19	540	1	216400
20	570	1	228400



Observe que as equações de regressão dos dois bairros, quando analisados separadamente são determinísticas, conduzindo a coeficiente de determinação R^2 igual a 1, ou seja, explicam a formação de preços com base nas variações das respectivas áreas, e apresentam resíduos para todos os elementos iguais a zero.

1ª. Tentativa: Tente juntar os dados dos dois bairros, numa só equação de regressão linear, utilizando duas variáveis explicativas: a variável área e a variável dummy, que assume o valor zero para o bairro 1, e o valor 1 para o bairro dois, conforme mostrado na tabela, como rotineiramente vinha sendo feito.

Processada a regressão nos programas especializados, inclusive no Excel, encontra-se uma equação de regressão com R^2 ajustado de 0,977, mas com erro padrão da regressão de 10997,73732 (desvio padrão dos resíduos), correspondentes à seguinte equação

$$Y = -80900 + 307,6923077 \cdot \text{área} + 121453,8462 \cdot \text{dummy}$$

Observe a tabela de resíduos abaixo calculados, correspondentes a essa equação, que apresenta erros percentuais nada desprezíveis de:

62% para o elemento n.1,

45% para o elemento n.2

39% para o elemento n. 10, e assim por diante.

Observe também que, para elementos próximos do centróide amostral, os resíduos são muito próximos de zero, como soe acontecer com as equações de regressão obtidas pelo método dos mínimos quadrados dos resíduos.

RESULTADOS DE RESÍDUOS						
<i>Observação</i>	valor total	valor calculado	Resíduos	Resíduos pad	erro porcentual	
1	30100	11407,69	18692,31	1,80	62%	
2	32100	17561,54	14538,46	1,40	45%	
3	34100	23715,38	10384,62	1,00	30%	
4	36100	29869,23	6230,77	0,60	17%	
5	38100	36023,08	2076,92	0,20	5%	
6	40100	42176,92	-2076,92	-0,20	-5%	
7	42100	48330,77	-6230,77	-0,60	-15%	
8	44100	54484,62	-10384,62	-1,00	-24%	
9	46100	60638,46	-14538,46	-1,40	-32%	
10	48100	66792,31	-18692,31	-1,80	-39%	
11	120400	132861,54	-12461,54	-1,20	-10%	
12	132400	142092,31	-9692,31	-0,93	-7%	
13	144400	151323,08	-6923,08	-0,67	-5%	
14	156400	160553,85	-4153,85	-0,40	-3%	
15	168400	169784,62	-1384,62	-0,13	-1%	
16	180400	179015,38	1384,62	0,13	1%	
17	192400	188246,15	4153,85	0,40	2%	
18	204400	197476,92	6923,08	0,67	3%	
19	216400	206707,69	9692,31	0,93	4%	
20	228400	215938,46	12461,54	1,20	5%	
valor total médio amostra		106750				
coeficiente de variação =		0,10				

2ª. Tentativa: Agora junte os dados dos dois bairros, numa só equação de regressão linear, utilizando três variáveis explicativas: a variável área, a variável dummy, que assume o valor zero para o bairro 1, e o valor 1 para o bairro dois, conforme mostrado na tabela e como segunda variável explicativa a variável de interação, resultante da multiplicação da variável dummy com a variável área.

Processada a regressão linear nos programas especializados, inclusive no Excel, encontra-se uma equação de regressão com R2 ajustado de 1, conduzindo todos os resíduos a zero, pois, como já se sabia, o modelo era determinístico.

6.2 - O PERIGO DOS CÓDIGOS ALOCADOS

Observe que na tabela de dados anteriormente utilizados, os dados do agrupamento bairro 1 correspondem exatamente a 4 vezes os dados do bairro 2, de forma determinística. Vamos então simular a situação em que o perito utiliza códigos alocados na observação dos dois bairros, ou seja, utilizando uma variável dummy que assume o valor 1 para o bairro 1 e o valor 4 para o bairro 2.

A tabela de dados para a regressão resulta:

EQUAÇÃO BAIRRO DE PESO 1 $y = 100 + 100 \cdot \text{ÁREA}$

EQUAÇÃO BAIRRO DE PESO 4 $y = 400 + 400 \cdot \text{ÁREA}$

ELEMENTO	área	BAIRRO	valor total
1	300	1	30100
2	320	1	32100
3	340	1	34100
4	360	1	36100
5	380	1	38100
6	400	1	40100
7	420	1	42100
8	440	1	44100
9	460	1	46100
10	480	1	48100
11	300	4	120400
12	330	4	132400
13	360	4	144400
14	390	4	156400
15	420	4	168400
16	450	4	180400
17	480	4	192400
18	510	4	204400
19	540	4	216400
20	570	4	228400

Faça a regressão linear, usando como variável explicada o valor total e como variáveis explicativas as variáveis área (quantitativa) e a variável qualitativa bairro, indicadas na tabela acima. Observe que embora o avaliador tenha “adivinhado” exatamente os pesos equivalentes aos dois agrupamentos de dados, a equação de regressão apresenta os mesmos resíduos apresentados na regressão acima utilizada, , embora com diferentes regressores.

Por último, introduza na equação acima detalhada mais uma variável explicativa, correspondente à variável de interação (multiplicação da variável “bairro”, com pesos 1 e 4, com a variável área) e observe como agora os resíduos são todos nulos.

Essa é uma explicação simplória do “porquê” na norma brasileira NBR14653-2 os códigos alocados serem considerados como a menos recomendável forma de montagem dos modelos, exigindo-se números mínimos de elementos amostrais em cada código, fato que no exemplo acima não é atendido.

São Paulo, agosto de 2005.