

Análise da Variável Data na Inferência Estatística

MARIA TERESA CERQUEIRA

Mestranda em Economia (UFBA)

Especialista em Avaliações e Perícias (UNIJORGE)

Especialista em Finanças Empresariais (UNIJORGE)

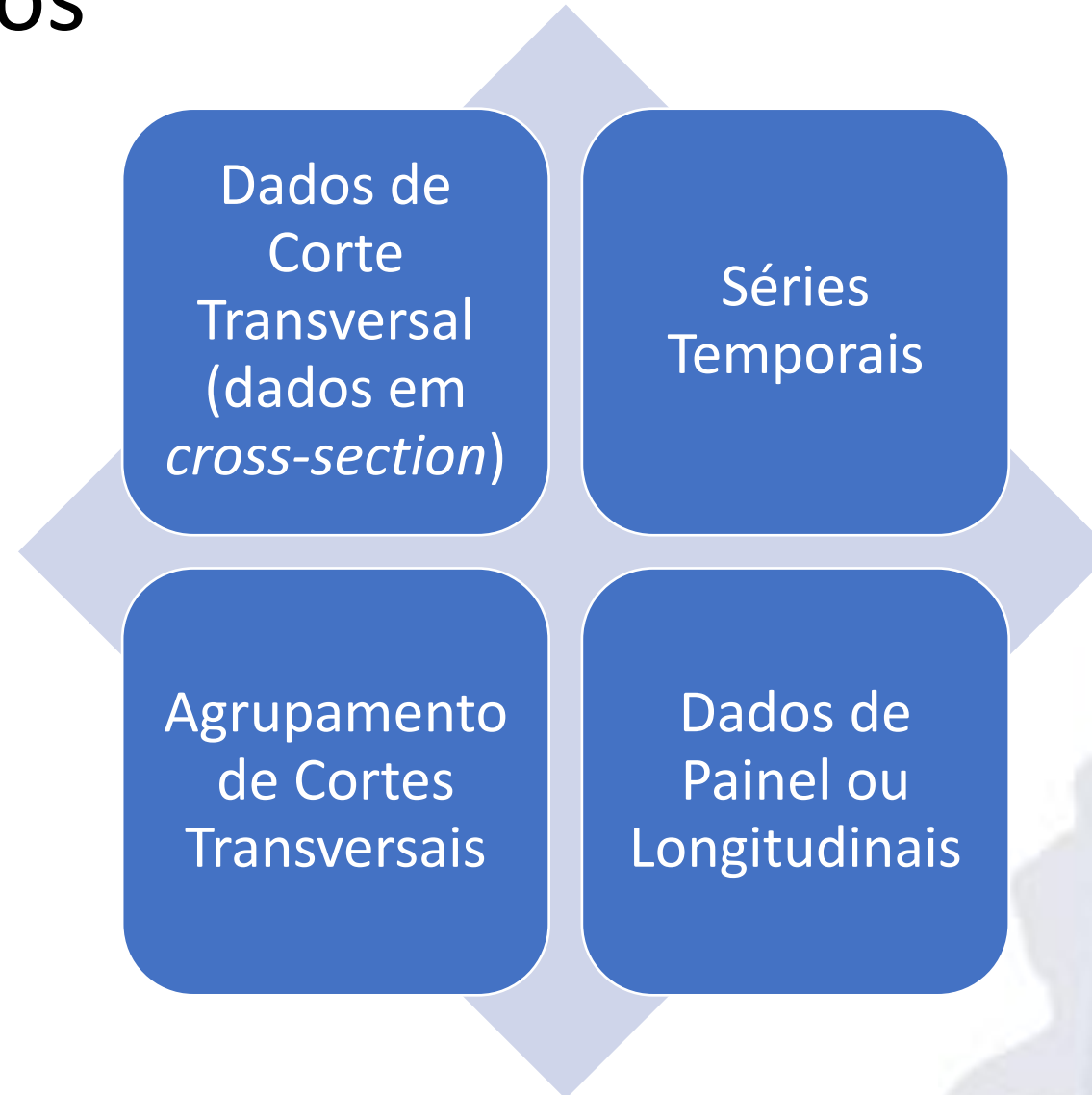
Engenheira Civil (UFBA)

Certificação IBAPE - Nível AAA

Vice-Presidente IBAPE/BA



Estrutura de Dados



REALIZAÇÃO



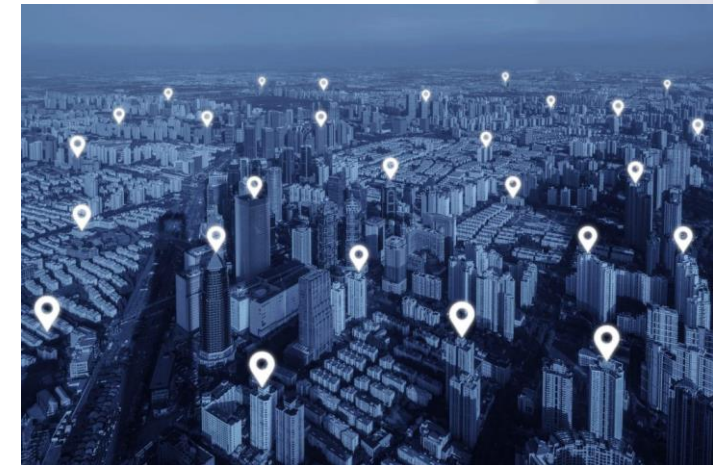
PATROCÍNIO



Estrutura de dados

✓ Dados de Corte Transversal (dados em *cross-section*)

- “Estes são dados em que uma ou mais variáveis foram coletadas no mesmo **ponto do tempo**”. (GUJARATI, 2006);
- “Consiste em uma amostra de indivíduos, consumidores, empresas, cidades, estados, países ou uma variedade de outras unidades, tomadas em um determinado **ponto do tempo**”. (WOODRIDGE, 2006);
- “Trabalha-se com dados em *cross-section* sempre que se tem observações para várias unidades individuais (e.g. pessoas, empresas, municípios, estados, países) em um **ponto no tempo**”. (MALBOUISSON; TIRYAKI, 2017);
- A palavra chave encontrada em todas as descrições aqui citadas é determinado “**ponto do tempo**”, ou seja, é o caso das Avaliações imobiliárias que se utilizam de uma amostra de dados coletada naquela **data base**.



Estrutura de dados

- ✓ **Dados de Corte Transversal (dados em *cross-section*)**

- “**fotos**” do mercado naquele instante;
- com uma base de **dados aleatórios**;
- É o caso das avaliações imobiliárias;

- **IMPORTANTE:**

Mesmo que a pesquisa não ocorra no mesmo dia ou semana, ignora-se quaisquer diferenças de tempo não importantes na análise de cortes transversal.

- **IMPORTANTE:**

A ordenação dos dados, a partir da amostra aleatória, não é relevante, ordenar os dados dentro do modelo por ordem alfabética, por bairros, por ordem crescente da área privativa, ou qualquer outra variável explicativa, não traria quaisquer prejuízo à modelagem.

REALIZAÇÃO



PATROCÍNIO



Estrutura de dados

✓ Séries Temporais

- “É um conjunto de observações dos valores que uma variável assume em **diferentes momentos do tempo**”. (GUJARATI, 2006);
- “Um modelo de série temporal reflete o padrão de movimentos passados de uma variável e usa essa informação para prever seus movimentos futuros”. (PINDYCK; RUBINFELD, 2004);

Exemplos de dados de séries temporais incluem preços de ações, oferta de moeda, índice de preços ao consumidor, Produto Interno bruto (PIB), indicadores macroeconômicos, taxas de ocupação de um resort, relatórios meteorológicos, taxa de desempregos, taxa de criminalidade, entre tantos outros.



Estrutura de dados

✓ Séries Temporais

- Um dos **objetivos** principais de uma análise de série temporal é a **possibilidade de previsões futuras** com base no comportamento passado dessa mesma variável;
- Nos modelos de regressão com base de dados de séries temporais, umas das **características** mais relevantes é que as **observações vizinhas são dependentes** e o modelo de regressão busca modelar esta dependência. Nestes modelos, a ordem das observações é de extrema importância;
- Esses dados são coletados em intervalos regulares, que pode ser diariamente, semanalmente, mensalmente, trimestralmente, anualmente, quinquenalmente (a cada cinco anos) ou decenalmente (censo demográfico). Existem ainda aqueles que são coletados em intervalos extremamente curtos, como exemplo das ações, que são obtidas em cotação de tempo quase real.

Estrutura de dados

✓ Agrupamento de Cortes Transversais

- Alguns conjuntos de dados têm tanto características de corte transversal quanto série de tempo;
- Este tipo de estrutura ocorre quando um mesmo conjunto de variáveis é coletado em diferentes períodos de tempo, em distintas amostras aleatórias de uma mesma população;

“Como por exemplo, considere o seguinte conjunto de dados sobre os preços da moradia coletados em 1993 e 1995 nos EUA, quando houve uma redução nos impostos sobre a propriedade em 1994. Suponha que tenhamos dados sobre 250 residências para 1993 e sobre 270 para 1995 e vamos tabular esses dados utilizando a uma nova variável (ano) separada. **Embora a ordem na qual armazenamos os dados não se revele crucial, não se esqueça de que o ano de cada observação é, geralmente, muito importante. Essa é a razão de introduzimos ano como uma variável separada.”** (WOOLDRIDGE, 2006).



Estrutura de dados

- ✓ **Agrupamento de Cortes Transversais – Quais os benefícios desse Agrupamento?**
- Com a utilização desse tipo de dados, além de **aumentar o tamanho da amostra**, a análise do corte transversal agrupada é importante para **estimar como uma relação fundamental mudou ao longo do tempo**;
- Podendo obter ainda, **estimadores mais precisos e estatísticas de testes mais poderosas**.



Estrutura de dados

✓ Dados em Painel

- “Consiste em uma **série de tempo para cada membro do corte transversal** do conjunto de dados”. (WOODRIDGE, 2006);
- “São estruturados com observações de variáveis de diferentes unidades em *cross-section* ao longo de um mesmo período de tempo, sendo que se tem **mais de uma observação temporal por unidade de *cross-section***”. (MALBOUISSON; TIRYAKI, 2017);

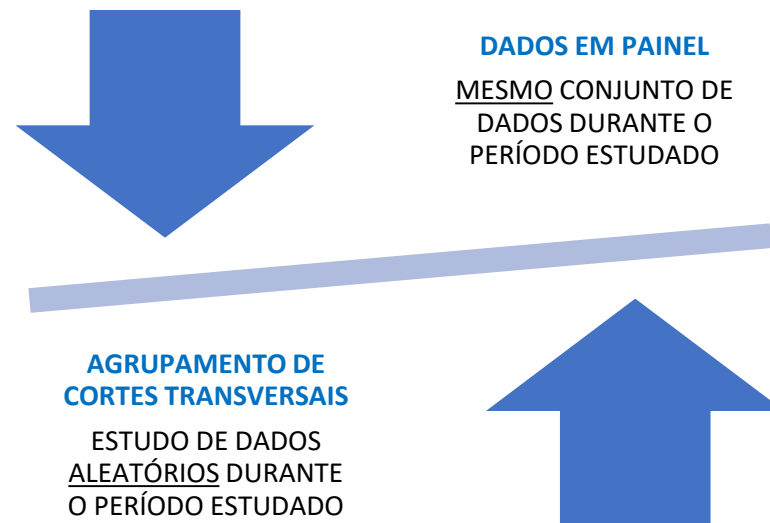
VANTAGENS:

- Frequentemente, permite estudar a importância das defasagens do comportamento ou resultado de tomar decisões;
- Podem ajudar a responder questões que outro tipo de estrutura de dados não permite.



Estrutura de dados

✓ Dados em Painel X Agrupamento de Dados



“A **característica essencial** dos dados de painel que os **distingue** dos dados de corte transversal agrupado é o **fato de que as mesmas unidades** de corte transversal são acompanhadas ao longo de um determinado período.” (WOOLDRIDGE, 2006).

REALIZAÇÃO



PATROCÍNIO



Qual estrutura de dados utilizamos nos macromodelos????

Diante das classificações das estruturas de dados apresentadas



- Verifica-se que os macromodelos de regressão que apresentam uso da variável data do evento se enquadram como **Cortes Transversais Agrupados**, visto que os dados coletados ao longo do tempo são aleatórios e que existe uma combinação entre períodos diferentes de vários cortes transversais;
- É bom ressaltar que nestes modelos **a ordem dos dados não se revela crucial**, no entanto, a determinação da sequência exata da variável data do evento é de extrema importância.

Pressupostos do Modelo de Regressão

Anexo A (normativo)

Procedimentos para a utilização de modelos de regressão linear

A.1 Introdução

A.1.1 A técnica mais utilizada quando se deseja estudar o comportamento de uma variável dependente em relação a outras que são responsáveis pela variabilidade observada nos preços é a análise de regressão.

A.1.2 No modelo linear para representar o mercado, a variável dependente é expressa por uma combinação linear das variáveis independentes, em escala original ou transformadas, e respectivas estimativas dos parâmetros populacionais, acrescida de erro aleatório, oriundo de:

- efeitos de variáveis não detectadas e de variáveis irrelevantes não incluídas no modelo;
- imperfeições acidentais de observação ou de medida;
- variações do comportamento humano, como habilidades diversas de negociação, desejos, necessidades, compulsões, caprichos, ansiedades, diferenças de poder aquisitivo, diferenças culturais, entre outros.

A.1.3 Com base em uma amostra extraída do mercado, os parâmetros populacionais são estimados por inferência estatística.

A.1.4 Na modelagem devem ser expostas as hipóteses relativas aos comportamentos das variáveis dependentes e independentes, com base no conhecimento que o engenheiro de avaliações tem a respeito do mercado, quando serão formuladas as hipóteses nula e alternativa para cada parâmetro.

A.2 Pressupostos básicos

Ressalta-se a necessidade, quando se usam modelos de regressão, de observar os seus pressupostos básicos, apresentados a seguir, principalmente no que concerne à sua especificação, normalidade, homocedasticidade, não-multicolinearidade, não-autocorrelação, independência e inexistência de pontos atípicos, com o objetivo de obter avaliações não tendenciosas, eficientes e consistentes, em especial as seguintes:

Recomenda-se que as características específicas do imóvel avaliando estejam contempladas na amostra utilizada em número representativo de dados de mercado;

- b) atentar para o equilíbrio da amostra, com dados bem distribuídos para cada variável no intervalo amostral;
- c) os erros são variáveis aleatórias com variância constante, ou seja, são homocedásticos;
- d) os erros são variáveis aleatórias com distribuição normal;
- e) os erros são não autocorrelacionados, isto é, são independentes sob a condição de normalidade;
- f) o engenheiro de avaliações deve se empenhar para que as variáveis importantes estejam incorporadas no modelo – inclusive as decorrentes de interação – e as variáveis irrelevantes não estejam presentes;⁸⁾
- g) em caso de correlação linear elevada entre quaisquer subconjuntos de variáveis independentes, isto é, multicolinearidade, deve-se examinar a coerência das características do imóvel avaliando com a estrutura de multicolinearidade inferida, vedada a utilização do modelo em caso de incoerência;
- h) não devem poder correlações evidentes entre o erro aleatório e as variáveis independentes do modelo, ou seja, o gráfico de resíduos não pode sugerir evidências de regularidade estatística com respeito às variáveis independentes;
- i) possíveis pontos influenciantes, ou aglomerados deles, devem ser investigados e sua retirada fica condicionada à apresentação de justificativas.

A.2.1.4 Verificação da autocorrelação

O exame da autocorrelação deve ser precedido pelo pré-ordenamento dos elementos amostrais, em relação aos valores ajustados e, se for o caso, às variáveis independentes possivelmente causadoras do problema.

Sua verificação pode ser feita, entre outros procedimentos, pela análise do gráfico dos resíduos cotejados com os valores ajustados, que deve apresentar pontos dispersos aleatoriamente, sem nenhum padrão definido.

REALIZAÇÃO

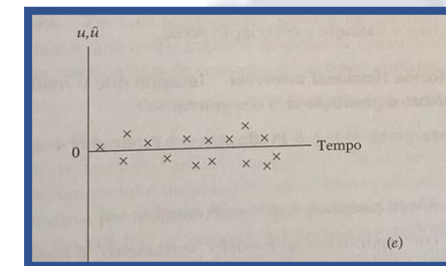
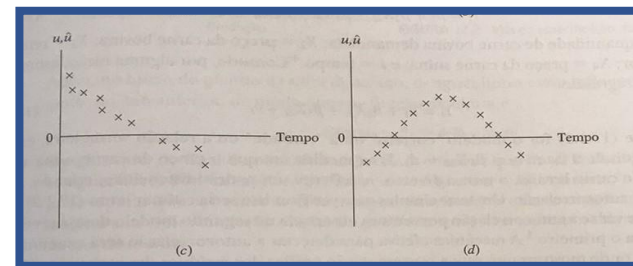
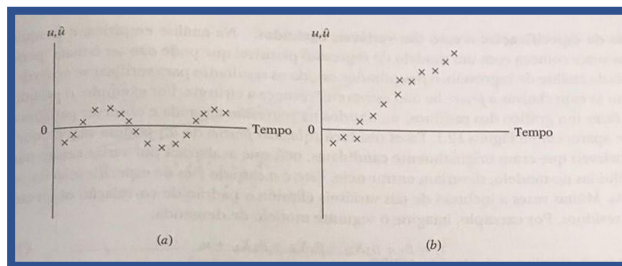


PATROCÍNIO



Autocorrelação

- O Modelo Clássico de regressão linear supõe que o termo de perturbação (erro) de uma observação qualquer não seja influenciado pelo de uma outra observação, ou seja, que o valor estimado da equação não esteja relacionado com qualquer outra previsão. **Quando um termo de erro em uma observação é correlacionado com o termo de erro de uma observação anterior, depara-se com o problema da autocorrelação;**
- “Correlações entre integrantes de séries de observações ordenadas no tempo, como as séries temporais, ou no espaço, como nos dados de corte transversais” (GUJARATI, 2006);
- Ou seja, a autocorrelação existe quando um termo de erro relacionado a qualquer das observações é influenciado pelo termo de erro de qualquer outra observação.



Fonte: Gujarati, 2000, pg 403

Autocorrelação

- Nas **estruturas de cortes transversais**, os dados são coletados através de amostras aleatórias, de forma que não há razões, **à priori**, para considerar que o termo de erro de uma observação aleatória teria qualquer correlação com o termo de erro de outra observação e quando esta autocorrelação é verificada, é denominada de **Autocorrelação espacial**, isto é, **correlação no espaço e não ao longo do tempo**;
- No entanto, quando se trabalha com **séries temporais**, os dados seguem um ordenamento natural ao longo do tempo e nestes casos, a pressuposição de que os erros correspondentes a observações diferentes **não estão correlacionados, não se mantém**. Sendo assim, nas séries temporais é muito difícil assumir que esta premissa seja atendida, **pois na grande maioria das vezes as séries estão fortemente relacionadas com seus históricos anteriores**;
- Existe ainda a possibilidade do que alguns autores denominam de **Correlação Serial**: Viés de especificação por Omissão de Variáveis Importantes ou por Forma Funcional Incorreta, e pode ser verificado no Gráfico de Resíduos x Valores Ajustados.



Autocorrelação x Mercado Imobiliário

- Dentro da Engenharia de Avaliações, a expectativa é que **cada transação se realize independentemente da outra**, ou seja, que os **dados de mercado sejam independentes**, de forma que o preço ofertado por um imóvel não interfira no outro;
- "O conceito de independência dos resíduos está ligado à independência dos dados de mercado. A situação ideal é aquela onde cada transação se realiza independentemente da outra. Isto é, o conhecimento do preço e condições de uma não interfira na outra."(DANTAS, 1998);
- "Desta forma, considerando as amostras do Mercado imobiliário colhidas em um mesmo período comportam-se como dados de corte transversal, podemos dizer que a presença de auto-regressão neste caso é pouco provável, pois dificilmente a perturbação ocorrida no preço médio de um determinado dado pode afetar os demais dados da amostra" (PELLI, 2014);
- O mercado imobiliário possui características de unidade geográfica, **com seus dados distribuídos num mapa** de acordo com a abrangência da pesquisa do avaliador, bem como apresenta **estrutura de dados de cortes transversais ou cortes transversais agrupados**. Sendo assim, se verifica, quando houver, a **autocorrelação espacial** e não a autocorrelação temporal, ou ainda a Correlação Serial (Erros de Especificação do Modelo).

REALIZAÇÃO



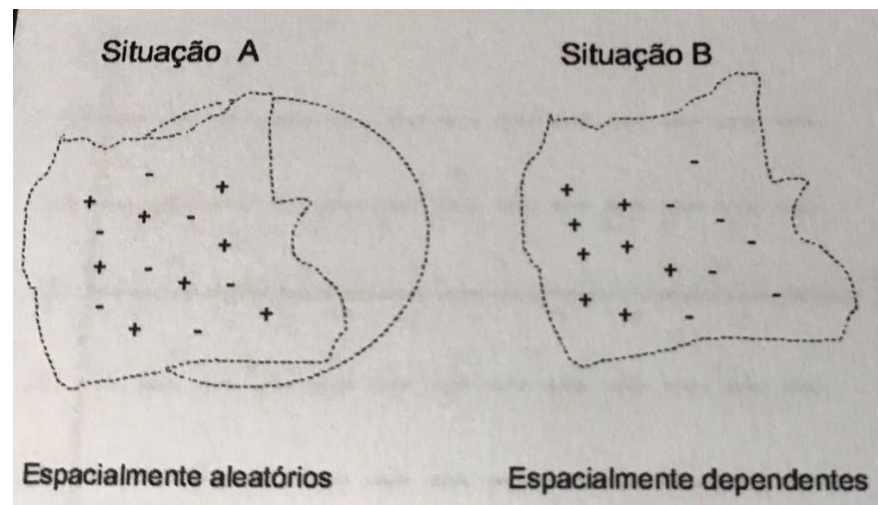
PATROCÍNIO



Autocorrelação Espacial

Autocorrelação espacial - a dependência espacial significa que a variável dependente (i) - numa determinada região, não é explicada apenas pelas suas variáveis explicativas, mas também pelo valor dessa variável dependente (j) – nas regiões vizinhas;

Autocorrelação espacial trata da possibilidade de apresentar interdependência entre as observações que são analisadas espacialmente, violando o pressuposto da esperança nula dos erros.



Fonte: DANTAS, 2013.

Autocorrelação Espacial

No mercado imobiliário **os objetos de análise são distintos entre si e possuem uma maior relação com a vizinhança próxima**, ou seja, os imóveis de um mesmo bairro tendem a apresentar condições semelhantes de localização e acessibilidade. A correlação espacial surge por falta de explicação correta, no modelo de regressão, das variações de preços no espaço.



REALIZAÇÃO



PATROCÍNIO



Autocorrelação Espacial

Anexo C
(informativo)

Recomendações para tratamento de dados por regressão espacial

C.1 Introdução

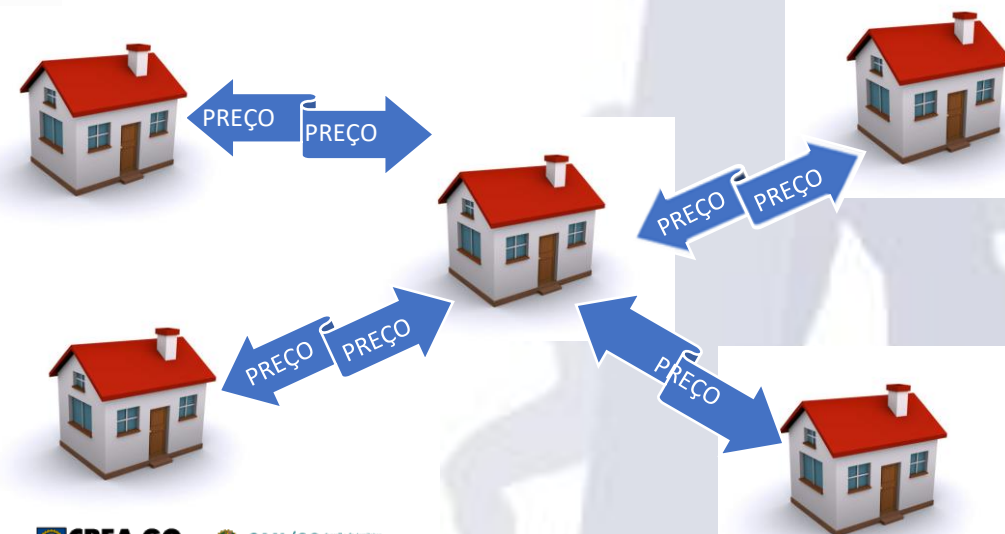
C.1.1 A regressão espacial é a técnica utilizada quando se deseja explicar a variabilidade observada em uma determinada variável dependente em relação às variáveis independentes, levando-se em conta a posição geográfica de cada uma das observações e as suas influências sobre os vizinhos.

C.1.2 Esta técnica é recomendada quando for constatada a existência de autocorrelação espacial entre os dados observados. O procedimento tem se mostrado especialmente útil em avaliações em massa, plantas de valores genéricos, estudos de velocidades de vendas e de demandas habitacionais, entre outros.

C.2 Pressupostos básicos

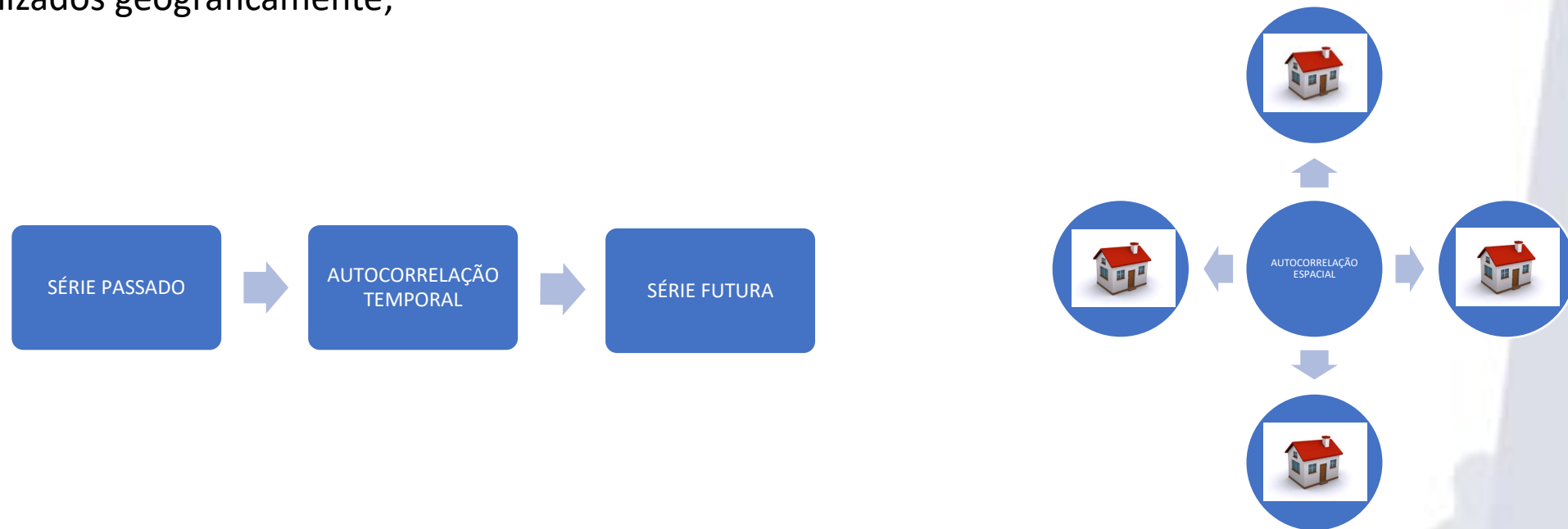
C.2.1 Devem ser observados todos os pressupostos da regressão linear clássica, expostos no Anexo A.

C.2.2 Caso seja verificada a existência de autocorrelação espacial, proveniente de interação ou dependência espacial entre os dados, recomenda-se incorporar os efeitos de dependência espacial ao modelo clássico de regressão, com o objetivo de assegurar as estimativas características de não tendenciosidade, eficiência e consistência.



Autocorrelação Espacial x Autocorrelação Temporal

É importante destacar a principal **diferença entre a autocorrelação temporal e autocorrelação espacial**, sendo que os dados organizados no tempo são relativamente mais fáceis de analisar do que os dados organizados geograficamente;

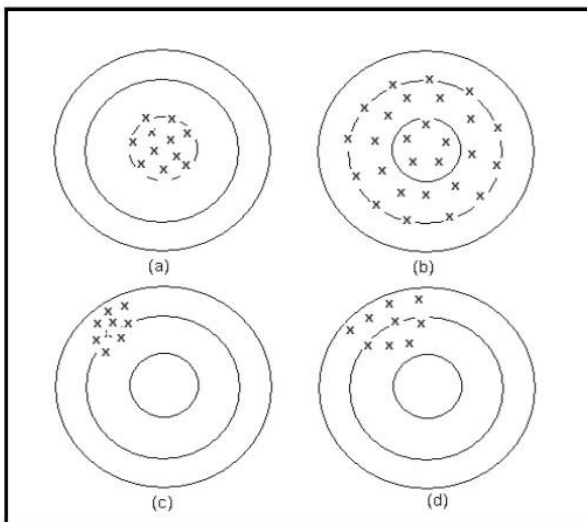


Ao contrário do espaço, o **tempo flui em um único sentido, unidirecional**, pois em séries temporais o passado pode influenciar o presente, mas o futuro, que ainda não aconteceu, não pode fazer o mesmo. Já no **espaço todas as direções, e todos os sentidos, podem ser relevantes**, dessa forma, a **dependência espacial possui multidireções**, ou seja, o valor de imóvel pode ser influenciado pelo seu vizinho a norte, a sul, a leste, a oeste, etc.

Propriedades Desejáveis dos Estimadores

Ausência de Tendenciosidade: Se diz que um estimador é não-tendencioso se a média ou o valor esperado de β^* for igual ao verdadeiro valor, ou seja, $E(\beta^*) = \beta$;

Consistência - O Estimador β^* é considerado consistente quando se aproxima do verdadeiro β conforme o tamanho da amostra aumenta. Ou, seja, espera-se que à medida que a amostra se torne muito grande, a probabilidade de que β^* seja diferente de β se torne muito pequena;



Fonte: IGM.MAT.BR, 2021.

Eficiência - Espera-se que um estimador seja não tendencioso eficiente quando, para um dado tamanho da amostra, a **variância de β^* seja menor que a variância de qualquer outro estimador não-tendencioso**. Algumas vezes é difícil julgar se um estimador é eficiente, então é natural descrever estimadores em termos de sua eficiência relativa, ou seja, um estimador será relativamente mais eficiente que outro.

REALIZAÇÃO



PATROCÍNIO



Violação do Pressuposto: Autocorrelação

- A presença da autocorrelação **interfere na eficiência** dos estimadores pelo MQO, ou seja, eles **deixam de ter variância mínima e assumem estimativas com variâncias amostrais desnecessariamente grandes**;
- E o erro padrão é viesado, o que conduz a intervalos de confiança e testes incorretos, como os testes t e F.

O Teste de Durbin Watson é adequado para analisar a presença da Autocorrelação da variável Data do Evento??



Um dos testes mais conhecidos para verificação de **autocorrelação temporal** é a estatística d , de **Durbin Watson**, que envolve o cálculo de um teste estatístico baseado nos resíduos do método de regressão de mínimos quadrados (MQO);

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

3.5.1 Análise do Teste de Durbin-Watson

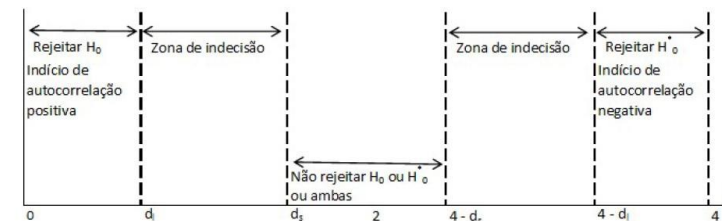
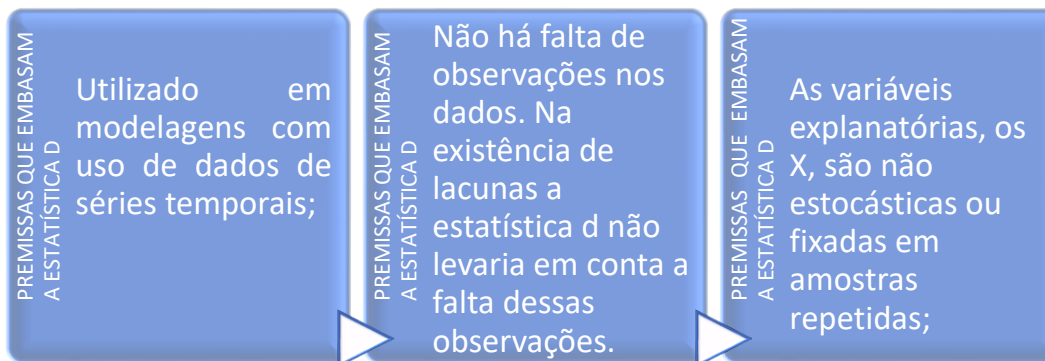


Gráfico 4 – Teste de Durbin Watson para autocorrelação dos resíduos

Sejam $\hat{e}$ ($t = 1, 2, 3, \dots, n$), os resíduos da regressão ajustada por Mínimos Quadrados, então se tem a razão entre a soma das diferenças, elevadas ao quadrado, entre sucessivos resíduos e a soma do quadrado dos resíduos (SQR).

Premissas para o uso do Teste de Durbin Watson

Verifica a hipótese de independência entre as observações versus a presença de correlação temporal



Em virtude dos modelos de regressão utilizados nas avaliações de imóveis pelo Método Comparativo com uso da **variável data do evento não trabalharem com a mesma amostra de dados, e sim com amostras aleatórias** coletadas em períodos diferentes do tempo, , **em um processo estocástico** e com uma **estrutura de dados em Agrupamento de Cortes Transversais**, bem como nos modelos de regressão utilizados em Avaliações, a ordem das observações é irrelevante para a análise. Ademais, a grande maioria dos modelos de regressão com uso desta variável **não possuem uma sequência ininterrupta na variável data do evento**, com a presença de várias lacunas de tempo;

O TESTE DE DURBIN WATSON SÓ DEVE SER UTILIZADO PARA SÉRIES TEMPORAIS, NÃO SE APLICA AS AVALIAÇÕES COMPARATIVAS DE DADOS DE MERCADO, MESMO NA PRESENÇA DA VARIÁVEL “DATA DO EVENTO”, POIS NÃO REPRESENTAM SÉRIES TEMPORAIS.

REALIZAÇÃO



PATROCÍNIO



Teste de Durbin Watson

1º. Modelo



Durbin-Watson statistic = 1

Região não conclusiva

Dados	x	Y
1	1	1
2	2	1
3	1	2
4	0	3
5	1	3

2º. Modelo



Durbin-Watson statistic = 1,5

Não autoregressão

Dados	x	Y
1	0	3
2	1	2
3	1	1
4	2	1
5	1	3

3º. Modelo



Durbin-Watson statistic = 2

Não autoregressão

Dados	x	Y
1	2	1
2	1	1
3	1	2
4	1	3
5	0	3

4º. Modelo



Durbin-Watson statistic = 3

Região não conclusiva

Dados	x	Y
1	0	3
2	1	2
3	1	1
4	1	3
5	2	1

5º. Modelo



Durbin-Watson statistic = 2

Não autoregressão

Dados	x	Y
1	0	3
2	1	1
3	1	2
4	1	3
5	2	1

O TESTE DE DURBIN WATSON NÃO SE APLICA ÀS AVALIAÇÕES COMPARATIVAS DE DADOS DE MERCADO, POIS NÃO REPRESENTAM SÉRIES TEMPORAIS.

Fonte: PELLI, 2014.

REALIZAÇÃO



PATROCÍNIO



Conclusão

- A estrutura de Dados utilizada nos modelos de regressão pelo Método Comparativo de Dados de Mercado é Dados em Corte Transversal (***Cross Section***) e **Agrupamento de Dados quando do uso da variável data do evento;**
- A presença da Autocorrelação, quando houver, é a **Autocorrelação Espacial ou Autocorrelação Serial (Erro de especificação por Omissão de Variáveis Importantes ou Má especificação da Forma Funcional);**
- **A presença da autocorrelação interfere na eficiência dos Estimadores, ou seja, eles deixam de ter variância mínima, trazendo prejuízo ao Modelo;**
- **O TESTE DE DURBIN WATSON É UTILIZADO PARA ANÁLISE DA AUTOCORRELAÇÃO TEMPORAL, NÃO SE APLICA AS AVALIAÇÕES COMPARATIVAS DE DADOS DE MERCADO.**

REFERÊNCIAS BIBLIOGRÁFICAS:

- ✓ ABNT, ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. Avaliação de bens. Parte 14653-1:2019.
- ✓ ALMEIDA, Eduardo, Econometria Espacial Aplicada. Campinas, SP, Editora Alínea, 2012.
- ✓ ARANHA, Francisco. Autocorrelação espacial na área de loja de supermercados nos municípios paulistas: mensuração por meio do Índice de Geary. Rev. adm. empres. [online]. 1999, vol.39, n.4, pp.38-45.
- ✓ CABRAL, Alexandra Maria Rios; LINS, Isis Didier. Aplicação de Bootstrap para teste de Durbin-Watson - Uma contribuição para cenários na economia. Economia política do desenvolvimento. Maceió, vol. 4, n. 12, p. 69-90, 2011.
- ✓ DANTAS, Rubens Alves. **Engenharia de Avaliações**: uma introdução à metodologia científica. São Paulo: Pini, 1998.
- ✓ GONZÁLEZ, Marco Aurélio Stumpf, Análise conceitual das dificuldades na determinação de modelos de formação de preços através de análise de regressão (2000).
- ✓ GUJARATI, Damodar N., Econometria Básica, Elsevier Editora Ltda, 2006.
- ✓ MALBOUISSON, Cláudia; TIRYAKI, Gisele F. Econometria na Prática, Alta Books Editora, 2017.
- ✓ MATTA, Túlio Alves. Avaliação do valor de imóveis por análise de regressão: estudo de caso para a cidade de Juiz de Fora. Juiz de Fora: Universidade Federal de Juiz de Fora, 2007.
- ✓ NETO, Antônio Pelli., Avaliação de Imóveis urbanos com a Utilização da variável data do evento e a estatística de Durbin Watson.
- ✓ Trabalho de Conclusão de Curso - TCC Maria Teresa Cerqueira (2017).
- ✓ TEIXEIRA, Marcelo da Costa; TEIXEIRA, Vanessa Domingos Gonçalves. Proposta de exclusão da variável data (ou tempo) em macromodelos estatísticos (inferência) com análise de estudo de Caso. Congresso Brasileiro de Engenharia de Avaliações e Perícias. Foz do Iguaçu. 2017.
- ✓ WOOLDRIDGE, Jeffrey M., Introdução a Econometria, Uma abordagem Moderna. Editora Cengage Learning, 4ª Edição, 2006.



REALIZAÇÃO



PATROCÍNIO



"A mente que se abre a uma nova ideia jamais voltará ao seu tamanho original" (Albert Einstein)

MUITO OBRIGADA!!!

Maria Teresa Cerqueira



@mariateresa.eng



@mariateresa.eng



@mariateresacerqueira



REALIZAÇÃO



PATROCÍNIO

